



Humboldt-Universität zu Berlin

**A Bibliometric Analysis of the Publications of the
Cluster of Excellence *Image Knowledge Gestaltung* in
the Context of Research Assessment**

Zur Erlangung des akademischen Grades Master of Arts

**Philosophische Fakultät I
Institut für Bibliotheks- und Informationswissenschaft**

Vorgelegt von: Tahsin Özdemir

Gutachter:

Prof. Dr. Michael Seadle

Alexander Struck, MA

Eingereicht am: 04.07.2016

Table of Contents

Table of Contents.....	I
List of Figures.....	III
List of Tables.....	IV
1. Introduction	1
2. Bibliometrics in Research assessment	4
2.1. Citations.....	7
2.2. H-index.....	8
3. Arguments against quantitative Indicators in Research Assessment	9
4. Data Collection and Processing.....	14
4.1 Selecting the Citation Databases.....	15
4.2. Queries and Variables.....	22
4.3 Data Cleansing	26
5. Bibliometric Analysis	31
5.1. Publications	32
5.1.1. Publication per year	32
5.1.2. Publications per Publication Format.....	33
5.1.3. Publications per Research Field.....	34
5.2. Coverage	36
5.2.1. Total Coverage.....	36
5.2.2. Coverage per year	37
5.2.3. Coverage of Research Field.....	38
5.2.4. Coverage per Publication Format	40
5.2.5. Coverage of Authors	41
5.3. Citation Analysis	42

5.3.1. Descriptive Data	43
5.3.2. Citation Rates	45
5.4. H-index Analysis	48
5.4.1. H-index GS.....	48
5.4.2. H-index WoS	49
5.5. Discussion	49
6. Conclusion.....	52
7. Bibliography	55
8. Attachment	60

List of Figures

Figure 1: Composition of disciplines in the publication list of the research center BWG.....	17
Figure 2: Proportion of research fields in the publication lists	18
Figure 3: Publication types covered in GS and WoS in a coverage study by Belew, R.	19
Figure 4: Proportion of publication formats in the publication lists	20
Figure 5: Profile of Peter Fratzl in Google Scholar (Screenshot)	28
Figure 6: Author identifier search option in Web of Science (screenshot).....	29
Figure 7: Exact-match-option using the <i>h</i> -index calculator in Google Scholar (Screenshot)..	30
Figure 8: Publication output per year	33
Figure 9: Proportions of publications formats in the data set	34
Figure 10: Publications per year splittet into the three main research categories	35
Figure 11: Pie chart visualizing the degree of coverage in Google Scholar	37
Figure 12: Publication covered in GS per year	38
Figure 13: Coverage of the publications distinguished by the three research categories	39
Figure 14: Coverage in GS per publication format	40
Figure 15: Boxplot of the citation variable	44
Figure 16: Citations (mean) per research field	45
Figure 17: Citations (mean) per year.....	46
Figure 18: Citations in total per year.....	46
Figure 19: Citation rate curves for the three research field categories.....	47
Figure 20: The mean value of the <i>h</i> -indices per research field for the data from GS	48
Figure 21: The mean value of the <i>h</i> -indices per research field for the data from WoS.....	49

List of Tables

Table 1: Publication output per year	32
Table 2: Publication per year distinguished by type of publication	33
Table 3: Publication per research field	35
Table 4 Coverage of the BWG publication data in Google Scholar	36
Table 5: Publication coverage in GS (total) per year	37
Table 6: Research fields of the BWG publications covered in Google Scholar.....	39
Table 7: Descriptive statistics with the transformed author variable	41
Table 8: Authors covered in Google Scholar	42
Table 9: Authors covered in Web of Science.....	42
Table 10: Descriptive statistics with the citation variable	43
Table 11: Citation numbers higher than zero (> 0).....	44

1. Introduction

This thesis presents bibliometric data of the Cluster of Excellence (Cluster) “Image Knowledge Gestaltung. An interdisciplinary Laboratory” (BWG)¹. In order to conduct a bibliometric analysis, this study attempts to measure the coverage of the publications of the Cluster in the citation databases Google Scholar (GS) and Web of Science (WoS). The question arises, if the publications of BWG are covered extensively enough in citation databases, to assess the research center with bibliometric indicators. The purpose of this thesis is to provide insights of the research performance of the Cluster by analyzing the publication data with bibliometric calculations.

In order to have a broader basis to answer the research question, information about the coverage of *h*-indices are provided for the respective authors in the publication lists in two citation databases Google Scholar and in Web of Science. Beside the coverage of *h*-indices, the actual values for every author in the publication lists will be used to calculate the *h*-indices with other variables in the data set.

As this thesis will show, I claim, that bibliometric indicators should not be used in an evaluation process of the Cluster of Excellence BWG with the expiration of the Excellence Initiative, but present arguments, when such quantitative indicators can be used to assess research performance. One reason for the negligence of bibliometric indicators is the potential unavailability of accurate bibliometric data in citation databases. I will not only support my thesis with bibliometric results of the research output, but also discuss the negative aspect of quantitative indicators in research assessment by citing critical sources regarding this active debate in research policy. There are debates about the applicability and the usefulness of bibliometric indicators in research assessment, especially in certain research areas (Bornmann 2015).

One crucial part of this thesis consists of a coverage study of the publications and the *h*-indices in the two citation databases Google Scholar and Web of Science, for an adequate coverage in academic databases is a precondition for the retrieval of bibliometric indicators. Therefore, this study answers the question to what extend the publications of this young research cluster are covered in Google Scholar, in order to be able to conduct bibliometric analyzes. Since it has become a standard in many research fields to enrich peer reviews with bibliometrics, the negligence of these indicators must follow proven and comprehensive logic.

¹ The German title of this Cluster is “Bild Wissen Gestaltung. Ein interdisziplinäres Labor”

These quantitative indicators, in this case bibliometric indicators enhance a peer review assessment to a so-called informed peer review (Jansen 2015). A general use of bibliometrics in research evaluation will be discussed in chapter 3.

For the extraction of bibliometric information, a citation database has to be chosen. In a following separate chapter, I will discuss why I used Google Scholar and Web of Science for the retrieval of bibliometric indicators. Since I have worked as a student assistant in the internal research library in the time during my master-thesis, I backed up my examination with the management and the editorial staff of the research cluster, especially with regard to the provided publication lists. The bibliometric analysis is based on the publication list of the editorial department. During my research, the management of the cluster taught me details and background knowledge about ongoing evaluation processes.²

First, I provide basic information about the young research cluster Image Knowledge Gestaltung in the light of the Excellence Initiative and the evaluation process. Then a discussion about the right measurement of research performance follows. Since there are several approaches to measure the impact and performance of research, the main measurements with quantitative and qualitative indicators are analyzed concerning the right evaluation method of an interdisciplinary research cluster.

After that, I will focus on bibliometric numbers and information using Google Scholar. In effect, there are some studies conducted, comparing the coverage and the extraction of bibliometric information of different citation databases, such as Thomson Reuter's Web of Science and Elsevier's Scopus (Falagas 2007). I will argue, why Google Scholar can be used for this investigation in contrast to the other databases.

Furthermore, when it comes to the researcher's *h*-indices, a comparative examination between the results of Google Scholar and Web of Science will be presented. I will show, that Google Scholar provides more data concerning the *h*-index. To calculate the *h*-index with GS data, I have used an *h*-index calculator, which is a software tool for Internet browsers (Ianni, G. et al. 2010).

In the following chapter, I describe in detail the methods and approaches for this examination. The two main sources of data are the publication list and the staff list of the research cluster.

² I have worked from March 2015 to May 2016 as a student assistant in the library of the research center BWG

One important factor of bibliometrics for research evaluation is the calculation of weighted citation rates. For this purpose, I calculate the citation rate with the retrieved data for each discipline for the given years. This must be done, in order to be able to compare the research performance of the different disciplines with each other.

“It is expected that science is at least likely to cite work in the distant past with the social sciences next, and the arts and humanities are expected to be most likely to cite from the older literature”

(Barnett 1999)

As this quotation by Barnett shows, a closer look at the citations per discipline is necessary to compare the different citation rates with each other. Also, with regard to the usefulness of bibliometric indicators for an evaluation, citation rates gain further weight, when considering the interdisciplinary composition of the research cluster and the relative short time frame of three years of publication data.

But beforehand, a general data evaluation will be provided with every variable of the calculation table, in order to understand the composition of the interdisciplinary research cluster and decide which citation database to use.

BWG is only one of four Cluster of Excellence of the Humboldt-University of Berlin (Humboldt Universität zu Berlin 2016), in the context of the Excellence Initiative. The Excellence Initiative is a German national research-funding project with the purpose to fund excellent research. The following quote taken from the website of the German national funding Organization Deutsche Forschungsgemeinschaft (DFG) explains the main principle behind this concept:

“The Excellence Initiative aims to promote top-level research and to improve the quality of German universities and research institutions in general, thus making Germany a more attractive research location, making it more internationally competitive and focusing attention on the outstanding achievements of Germany universities and the German scientific community”

(Deutsche Forschungsgemeinschaft 2016a)

An important part of Excellence Initiative is the so-called Clusters of Excellence (in German: Exzellenzcluster). The DFG further explains the concept Cluster of Excellence:

“Clusters of Excellence will enable German university locations to establish internationally visible, competitive research and training facilities, thereby enhancing scientific networking and cooperation among the participating institutions. Clusters of Excellence should form an important part of a university's strategic and thematic planning, significantly raise its profile and reflect its considered long-term priorities.

They should also create excellent training and career conditions for young researchers. In conjunction with the other two funding lines, i.e. graduate schools and institutional strategies to promote top-level research, clusters of excellence will help to increase Germany's attraction as a research location in the long term and improve its international competitiveness”

(Deutsche Forschungsgemeinschaft 2016b)

Basis of this examination are the publication lists of the Cluster BWG from 2012/13 to 2015 for a bibliometric analysis. Since this research center has been established in November 2012 the first publication list comprises 2012 and 2013.

The main tables and figures are incorporated in this examination. Additional tables are attached to this work. All calculation tables used for this study are provided on external medium.

2. Bibliometrics in Research assessment

In this chapter, I consult arguments from different scholars concerning the use of bibliometric indicators for research assessment. Also, the benefits of citation databases are discussed, in the light of reliability and coverage of publication data for evaluations.

“Governments need systematic evaluations for optimizing their research allocations, reorienting their research support, rationalizing research organizations, restructuring research in particular fields, or augmenting research productivity” (Moed 2009)

This quote by Moed indicates the importance of research evaluations for the allocation of research funds, with respect to the effectiveness of research policy. The question arises, which method to use, in order to get the best results for research assessment.

Since there is no standard for the evaluation of research success with quantitative indicators in Germany, but rather recommendations by funding organizations, such as the Wissenschaftsrat (WR), each research center is evaluated differently. Only the research institutes of the Max Planck Society do have a system “partly based on metrics” (Abbott 2010), but still there is no standard system for the measurement of research performance.

There is an ongoing debate about the use of quantitative indicators in Germany, concerning the measurement of research performance. Germany, unlike other nations in Europe, focuses on the amount of third-party funds as the main quantitative indicator in research assessment. Whereas other countries, mostly from correspondent national funding organizations, use this indicator as

an additional metric, in Germany the third-party fund dominates the debate about quantitative indicators in research assessment (Jansen 2015).

However, there are more and more experts, who strongly recommend using bibliometric indicators, in order to measure research performance. For example, Jürgen Gerhards, who works for the Berlin Brandenburgische Akademie der Wissenschaften (BBAW), recommends using more bibliometrics instead of the third-party fund indicator, because it would redress the unjustifiable allocation of funds for different research disciplines (Gerhards 2013).

Bibliometrics have become an increasing method to measure research quality. In natural sciences for example, it has become a widespread standard and even in the humanities bibliometric starts to gain greater importance, in particular with respect to budget cutting by the funding organizations. However, in Arts and the humanities bibliometric is not as common as in the life- and natural sciences (Bornman 2015).

“Die meisten international vergleichenden Evaluationsverfahren beziehen sich entsprechend in erster Linie auf Publikationen als zentraler Einheit zur Messung von Forschungsleistungen. Und sie orientieren sich an der Resonanz von Publikationen, wie sie sich in Zitationen manifestieren” (Gerhards 2013).

Still the quotation above shows, that most of the international evaluation systems in research policy relate to the measurement of publications and their citations. The only alternative, when avoiding all quantitative indicators, also bibliometrics, in research assessment would be the peer review evaluation. In fact, peer review is “the most important instrument for assessing scientific work” (Bornmann 2008).

But also, this traditional evaluation method has disadvantages. Opponents of peer review evaluations often point out to the “complex and expensive and subjective qualitative peer review evaluation processes” (Kostoff 1998).

That is one reason, why quantitative indicators gradually get more awareness in research assessment. In Germany, however, the main quantitative indicator is still the third-party-fund-indicator. An often-cited argument for the use of the third-party fund indicator is the fact that project applications must go through an evaluative and critical process until the funding organizations accept them. Nevertheless, bibliometric indicators gain more and more awareness, also in Germany and slowly also in research fields of the arts and humanities (Bornmann et al. 2015).

The question arises, which metric should be used to measure research performance and impact. Whereas publication productivity can simply be measured by publication counts, the so-called *impact* of publications is far more elaborate to measure. The following quote by Moed, introduces the most common method in bibliometrics, namely the citation analysis:

“Citation analysis involves the construction and application of a series of indicators of the “impact”, “influence”, or “quality” of scholarly work, derived from citation data, i.e. data on references cited in footnotes or bibliographies of scholarly research publications.” (Moed 2009)

As this quote explains bibliometricians analyze the citations of publications by building citation-based indicators and evaluated these numbers with statistical calculations. The motivation is to provide bibliometric indicators to enhance a peer review evaluation to a so-called informed peer review (Ochsner et al. 2014).

Also, the German Wissenschaftsrat (WR) recommends using informative peer reviews for research field from the humanities and arts:

“Dieses informed peer review ist sowohl einer reinen Reputationsbewertung als auch einer automatisierten indikatorenbasierten Bewertung vorzuziehen, da einerseits durch die Indikatorenbasis rein subjektive (Vor-) Urteile auf den Prüfstand gestellt werden und andererseits eine Einordnung und mögliche Relativierung der für sich genommen nicht immer hinreichend aussagekräftigen Indikatoren durch Fachleute stattfindet.”

(Wissenschaftsrat 2011)

This is in fact the main issue of concern, because it is necessary to find the right evaluation method. As Moed (2009) states, “the future of research assessment lies in the intelligent mixture of metrics (including bibliometric indicators) and peer review.” He goes on to claim, that policy makers, committees, who are responsible for peer reviews and bibliometricians should work together, in order to develop new methods for research assessment, which combine qualitative and quantitative indicators (Moed 2009). The following citation by Bornmann criticizes the focus on one single indicator and stresses the importance of the different aspects of performance and impact in research assessment:

“Several indicators are necessary in order to illuminate different aspects of performance and to provide a more adequate and multifaceted picture of reality” (Bornmann 2007).

In fact, for the assessment of research performance of this relative young Cluster of Excellence BWG, the right mixture must be found to ensure an evaluation, which addresses the interdisci-

plinary character of this research center. When we consider the strong representation of disciplines from the arts and humanities and the sciences in the interdisciplinary composition of the Cluster, the question arises, which assessment method will be best to measure the impact of the research output. This aspect is not only one subject of this thesis, but will also be discussed critically in the next chapter, together with a critical view on quantitative indicators in research assessment.

But first, the main metrics retrieved from the citation databases Google Scholar and Web of Science should be introduced, since they belong to the most important bibliometric indicators for the measurement of research impact.

2.1. Citations

Since the start of bibliometrics as a research method, citations counts are the most important metrics. Citation count is the basis for a number of citation based-indicators, like citation rates or the *h*-index (Ball and Tunger 2005). In evaluative bibliometrics they are the foremost indicators for research assessment (Wang 2013). Even, as mentioned, the excessive use of this indicator and its validness for measuring the impact of academic publications is controversially debated among experts (De Bellis 2009). For example, as I will discuss in the next chapter, the context of citations is not always clear or because the numbers must be taken from citation databases, which must have the publication covered first, before they can be used by experts and bibliometricians. Nevertheless, the counting of citations is a simple metric from the references in the bibliographies or foot- and endnotes.

Of course, it is not adequate to compare the citation counts from two total different research field with each other, since the citation pattern and publication behavior of different research disciplines can vary distinctively. Therefore, different impact factors are created with the citation count indicator. The aim is to have indicators for the evaluation of many authors, journals, publications and research fields (Havemann 2009; Moed 2009).

In this examination, I take the citation numbers from GS and calculate the citation rates of the three main broad research categories (see chapter 4). Another goal is to examine the applicability of the citation data due to the amount of data covered in the database.

Another important aspect about using citation rates to evaluate research success lies in the different citation and publication behavior of different disciplines. Given that a publication needs

a special time window to get cited, there are bibliometric studies, which show, that for example publications from the natural sciences have a shorter citation age than publication from the humanities (Barnett et al 1999; Wang 2013).

When we consider the interdisciplinary composition of the research cluster with a strong emphasis on the arts and humanities, a fast and only output-focused evaluation would be a wrong attitude toward a discipline, which needs more time to get visible in the scientific communication.

2.2. H-index

Another indicator used in this examination is the so-called *h*-index, developed by Jorge Hirsch in 2005 (Jacso 2008a). This number is an author-based indicator, measuring the impact of an author's publication output.

“The *h* index is seen to have the advantage that it gives a robust estimate of the broad impact of a scientist's cumulative research contributions” (Hirsch, 2005)

One advantage of the *h*-index is the simplicity of its calculation. To determine the *h*-index of an author, all of his or her publications must be ranked by citation counts. Then the *h*-index is the number, where the ranking position and the amount of citations are equal or below the ranking position, or as Hirsch puts it in his famous work:

„A scientist has index *h* if *h* of his or her *N_p* papers have at least *h* citations each and the other (*N_p h*) papers have *h* citations each. “(Hirsch 2005)

H-indices are mostly available in citation databases like Web of Science or Scopus (Elsevier 2016a). Since, it is quite easy to calculate the *h*-index, compared to other bibliometric indicators, it is also possible to determine this value manually, presupposed all publications with the amount of references are at hand. But in common practice the *h*-index is taken from citation databases.

In this thesis, the *h*-indices of the authors of the publications of the research center are gained from Google Scholar and Web of Science. In the next chapter I will outline the weaknesses of this indicator together with a general critique of the use of bibliometric indicators in the evaluation of the Cluster of Excellence BWG.

3. Arguments against quantitative Indicators in Research Assessment

This chapter provides arguments against the excessive use of bibliometric indicators in research assessment. First, there is a general skepticism about the informative value of the common indicators in bibliometrics. Even the significance of the standard bibliometric numbers, like citation counts are often debated in the discussion about research assessment (Kostoff 1998), (Abbott 2010). Another key-indicator, the *h*-index is even more criticized even among bibliometricians (Waltman 2012).

Because of the different publications and citation behavior of each research field, citation counts need to be normalized, in order to be able to compare the impact of publications from different disciplines. This leads to the so-called citation rates. In this study, I calculate the citation rate based on three main research categories by Thomson Reuters Science Citation Indices (Thomson Reuter 2016a). The database provider categorizes research disciplines for the covered journals in so-called “Citation Indices”. In the three main categories science, social science and arts and humanities, all relevant research disciplines are listed. For the calculation of the citation rates, I used the definition of citation rates on the company’s official Website:

“A citation rate is the average number of citations received by a group of papers published in one research field in a given year. It is calculated by dividing the total number of citations received by all papers in the group (defined by research field and publication year) by the total number of papers in the group” (Thomson Reuter 2016b).

In many studies, bibliometricians have shown, that the citation rates for publication from the humanities are lower than in other disciplines:

“Fields differ in citation patterns; work that is fundamentally of a historical character will necessarily have citations to primary sources that may be quite old” (Barnett 1999).

This means, that every publication needs a certain time until the publication becomes visible in the scientific communication and develop a stronger impact, measured in citations. The peak of citations is reached later in the arts and humanities compared with the other fields. That means bibliometric indicators in research assessments can’t be used immediately after the publication time. This leads Bornmann to claim: “When deciding which years of an institute’s publication record should be included in the analysis, care should be taken that the majority of the papers (if not all) are at least two years old”. (Bornmann et al. 2014 p. 204)

And Wang states: "...there are significant differences in citation ageing between different research fields. For studies on one specific field, a tailored citation time window is preferred...a 3-year time window is sufficient for the biomedical research fields and multidisciplinary sciences, while a 7-year time window is required for the humanities and mathematics." (Wang 2013)

The German national science organization, der "Wissenschaftsrat" (WR) advises in its recommendations for research assessment, that research center should be assessed with a frequency of five or ten years. This does not count explicitly for bibliometric evaluations, but impact of research output in the form of academic publications also need a certain time frame, in order to be measurable.

"Bei regelmäßig stattfindenden institutionellen Evaluationen sollten diese – gleich ob im außeruniversitären oder universitären Bereich – in der Regel nur alle fünf bis zehn Jahre durchgeführt werden. Bei der Festlegung der Intervalle sollte die Berufungsdauer des Leitungspersonals berücksichtigt werden. Durch größere Evaluationsintervalle wird der Aufwand für die Begutachteten sowie die begutachtenden Personen und Institutionen begrenzt, zudem werden Momentaufnahmen vermieden, mittelfristige Tendenzen sichtbar, und die zwischen zwei Evaluationen erfolgten Veränderungen können sinnvoll bewertet werden."

(Wissenschaftsrat 2011)

Otherwise important trends and developments in the research activities can be missed, beyond the decisive aspect of visibility of the research publications measured with citations.

This leads us to the fact that the Excellencecluster BWG is only shortly older than three years, which according to the recommendations of the highest entity would be too young to be evaluated. Together with the Bornmann statement above, which says that all publications must be at least 2 years old and given that a great share of publications have been published in 2015 an evaluation with publication data would have only minor explanatory power.

Another critical aspect of neglecting bibliometric indicators is the claim by some researchers, like Richard Münz, that the focus on quantitative indicators can hinder creative and innovative research work, because, especially when relating too much on the third-party fund indicator.

"Die Konsequenz einer auf die Erfüllung von Positionen und Reputation ist jedoch die Eindämmung der Vielfalt von Wissen und die Behinderung der Kreativität von Forschung." (Münz 2012).

Finally, an evaluation based on quantitative bibliometric indicators can only be used, when there is a sufficient degree of coverage in citation databases. This argument is the subject of examination in this thesis. Publication data must be queried in databases for the retrieval of bibliometric indicators. A research evaluation is always limited by the time frame of the investigation. In this case, the time frame starting with the establishment of the Cluster of Excellence in late 2012 (November 2012) until the end of 2015. A significant share of publications in the bibliographies of these years must be covered in citation databases in order to create comprehensive indicators. A complete coverage of bibliographies is in most cases improbable, but on the other hand reviewers and bibliometricians expects a decisive part of coverage, as this quote by Bornmann indicates:

“The meaningfulness of bibliometric data for research evaluation ultimately depends on the coverage of the publications in the databases selected. What is not covered by the databases can also not be evaluated.” (Bornmann 2015)

This leads to the question of a determined degree of coverage. Since there is no official standard, every evaluation based on bibliometric indicators must be transparent concerning the data sources.

When calculating the coverage of publications in a citation database, the question arises, to which extend the coverage of publications in a database has to be, in order to be useful for the assessment of research facilities with bibliometric indicators. I state, that a bibliometric analysis of the publication output of a research facility will be too limited if not even half of the data can be extracted from databases. When there is only a short percentage of the publication covered in the database the creation of indicators would be impossible or at least almost meaningless for the purpose of measuring the impact of the research publications. In this case, the bibliometric analysis will be restricted to a general publication count analysis. However, the examination of publication coverage in the respective citation databases can also be informative for a research assessment.

Information from citation databases can be used to build multiple indicators. The total coverage of publications indexed in citation databases is only the precondition of a comprehensive evaluation enriched with bibliometric numbers. On the other hand, one can argue that also the coverage of research publications in citation database can be seen as indicator of research quality, because of the potential extraction of bibliometric numbers.

Therefore, the degree of coverage necessary to conduct a bibliometric analysis needs to be defined. As mentioned above, I state that bibliometric indicators should not be included in the research assessment of the research center BWG, when not half of the publications are covered in a citation database.

Beside citation counts Google Scholar and Web of Science are contributors to data and metrics, that can be used not only to evaluate papers and journals, but also a whole research facility and every scientist in particular with the help of the *h*-index. The coverage of *h*-indices must also be considered for an evaluation.

The publication data must be accordingly represented in databases, in order to build main indicators, which can further be processed to new variables. Using data from databases, which only comprises approximately a third of the publication production of a research center, would not be very useful. The reasons for the lack of information from the database is in this case the main point of concern. For, the goals of research evaluations have all in common, that they seek visibility and impact in the scientific communication.

As mentioned, this examination is intended to conduct a bibliometric analysis and looks at the coverage of bibliometric indicators for the use in a full-fledged evaluation of a young research cluster of the Humboldt-University of Berlin. Because of the perspective from a national research program, which funds “excellent” research, the reasons for ignoring bibliometric numbers in an evaluation must be clear. Additionally, the avoidance of efforts to retrieve, process and analyze publication data of a research center can direct the focus to qualitative indicators and aspects of research activities.

In common use, bibliometric numbers are not used solely for the decision-making, but instead added to peer review reports. In this sense, a need for a complete coverage of publications in databases is not even necessary. But the bigger the share of publications from the bibliographies covered in databases the more the data contributes to valid statements.

A further aspect concerning the omission of bibliometric indicators in the context of citation databases is the language of the publications. Together with the research field, these two aspects influence the probability of coverage in citation databases, since it is known that for example ISI-based database only covers a small percentage of publication in German language (Bornmann et al. 2015).

Also, books, one main publication format in the humanities, are also not yet adequately covered in databases, although there are current intentions to build the so-called “Book Citation Index”

by Thomson Reuter (Thomson Reuter 2016c). Because citation analyses are mostly conducted with journals as publications, the decision to conduct a bibliometric analysis must consider the dominant publication format of the research output. In chapter 4, I present information about the exact representation of the amount of disciplines in the publications lists and show that journals only present a third of the total publication format. This is why a strong focus on bibliometric indicator would not be advisable, due to the expected limitation of available data.

However, quantitative indicators, in particular bibliometric indicators also have general disadvantages. When it comes to solely counting the citations of publications, the meaningfulness of the numbers must be questioned for many reasons. For example, citations don't tell in what context the publications had been cited (Stellenbosch University 2014). It must also be considered, that negative reasons of citations cannot be excluded from the numbers. Also, there is no distinction between citations, which are crucial to the citing publications and others, which are more irrelevant to the main aspects of the publications. This lack of weighted citations, leads to the question of normalization of the numbers. Since there are differences in citations patterns and publication behaviors in the different research disciplines, these factors must be taken into account in evaluative bibliometrics.

Another weakness of citation counts, as basis for an evaluation, is the common use of self-citations (Jacso 2008b). Nevertheless, as Hirsch states the effect of the h -index is much lesser than on the citation count indicator:

“Finally, in any measure of citations, ideally one would like to eliminate the self-citations. Although self-citations can obviously increase a scientist's h , their effect on h is much smaller than on the total citation count.” (Hirsch 2015)

One main aspect of concern in this study, regarding the use of the h -index from Google Scholar and Web of Science, is the fact that the h -index has a bias towards already established and older authors:

“The H-index, an appealing construct, is highly biased towards “older” researchers with long careers and those active in fields with high citation frequencies and provides an incomplete picture of a group's actual citation impact.” (Moed 2009)

For example, the h -index of one authors increasing over time, also without publishing any further research work (Glänzel 2006), (Waltman 2012). As it is known, that the given object of examination, namely the Cluster of Excellence BWG consists of many young scientists, high

h-indices should not be expected. Another weakness of the *h*-index is that because of the mentioned differences in citation patterns of the disciplines only authors from the same research field and also with the similar age can be compared with each other (Marx 2015).

This is why these indicators should be used with care, if at all, in the research assessment of this Cluster. Nevertheless, a bibliometric analysis will be conducted in this examination to prove this thesis.

4. Data Collection and Processing

In this chapter I delineate the approach and methodology for the data collection. Therefore, two factors are indispensable for the approach. First, this study examines the coverage of the publication data in citation databases. Second, the aim is to use the extracted bibliometric indicators to evaluate the suitability of these indicators in research assessment.

The first approach for the retrieval of bibliometric indicators for a potential evaluation is the selection of the right data source. Nowadays there are several providers of academic databases. Common citation databases are Elsevier's Scopus, Google Scholar and to Thomson Reuter's Web of Science. After selecting the database for the collection of bibliometric information, queries must be conducted on the basis of the publication list. The aim is to create a new variable and set the values accordingly. Again, we can separate two different kinds of queries in order to retrieve the two main bibliometric indicators, namely citation counts and *h*-indices. The title query is used for measuring the coverage of publication in Google Scholar and using the citation numbers for a bibliometric analysis. Author names are queried in both databases Web of Science and Google Scholar for the calculation of the *h*-indices.

Considering the coverage study, all academic publication titles in the bibliographies from 2012/2013 to 2015 were queried in Google Scholar. In this case a new variable in the calculation table must be created with two values representing the positive and negative result of the query. The query should incorporate the full title, in order to avoid similar matches. The identification of the entry is the main point of concern, so that all metadata of the entry must be checked, before setting the value in the variable accordingly. But beforehand, an explanation follows, why I used Google Scholar and Web of Science as data sources for a bibliometric analysis.

4.1 Selecting the Citation Databases

In the following, I present arguments, why Google Scholar is the most potential candidate, in terms of providing bibliometric data of the publication of the Cluster, in order to measure the coverage and to conduct a citation analysis. As mentioned above, the selection of an appropriate citation database also concerns the *h*-index. For this purpose, not only is Google Scholar used to calculate the *h*-indices of all authors from the publication lists, but the same author names were also queried in Web of Science, in order to determine which of the both databases provide a more comprehensive degree of coverage of the *h*-index.

To find out which database would provide the most data for the bibliometric analysis, it is necessary to get an overview of the research publication of the Cluster. One aspect of the publication policy of this young research center is the recommendation to publish Open Access (OA) (Bild Wissen Gestaltung 2016a). When considering, that Web of Science does not cover OA-publications broadly (Falagas et al 2007), it should be excluded as the source for the retrieval of bibliometric indicators of the research publications of the Cluster. According to McVeigh (2007) OA – Journals make up “approximately 1% of the 20,000 journals in ISI Web of Knowledge”. Further the author reports, that only highly cited OA-Journals are covered in WoS.

On the other side, there are numerous studies about the coverage of Open Access literature in Google Scholar, because Google does not give details about the coverage numbers (Orduna-Malea et al 2014). Also Jamali (2015) claims, that GS “plays a major role in finding free full-text versions of articles”.

However, in its coverage policy GS claims that it “includes journal and conference papers, theses and dissertations, academic books, pre-prints, abstracts, technical reports and other scholarly literature from all broad areas of research ... a wide variety of academic publishers, professional societies and university repositories, as well as scholarly articles available anywhere across the web. Google Scholar also includes court opinions and patents” (Google 2014)

In this sense it can be concluded, that GS covers more open OA articles, than WoS. Google Scholar was launched in 2007 and has gained more awareness by researchers since then, because of its fast-growing collection of academic bibliographic data. Another reason to use GS for bibliometric analyses is the free availability of the citation database.

Nevertheless, in the following I provide a concise introduction of the citation database Web of Science, formally known as ISI (Ball and Tunger 2005), because of its importance in bibliometrics and due the fact that a coverage study of the *h*-index is also subject of this thesis.

Traditionally, bibliometric calculations have been done with the bibliographic data of the Institute for Scientific Information (ISI), which was founded by Eugene Garfield in 1960 (Thomson Reuters 2008d)

“The ISI has historically been the exclusive provider of citation analysis data. However, other providers of scholarly analysis of citations have emerged. Scopus (Elsevier B.V., New York, NY) and Google Scholar (Google, Inc, Mountain View, CA) have challenged ISI as the only source of citation data.” (Thompson et al. 2009)

ISI was then the only reliably source for academic bibliometric information. Now there are more databases for academic publications. The ISI-based Web of Science by Thomson Reuters is still one of the most used for bibliometric studies (Moed 2009).

However, there are other citation databases now competing with Web of Science, as the quotation above by Thompson suggests. One commonly used database is Elsevier’s Scopus, which was introduced in 2004 (Elsevier 2016b).

This means, before the selection of the database, the research fields of the publications have to be analyzed. Thompson (2009) goes on to state that: “...the same search done on all 3 databases would probably generate 3 different results from the available citation networks. This must be kept in mind when evaluating faculty citation counts using different sources”.

This means, before a bibliometric analysis can be done, the right database must be chosen, to get the most useful result from the bibliometric evaluations. The expectation, that one citation database would have all data covered necessary for the analysis is on the other hand also unrealistic and improbable, in particular, when we consider the fact, that a significant share of the given publications are recent publications from 2015 and from the arts and humanities.

Therefore, a first look on the publication list of the research facility and on the composition of the interdisciplinary research projects are necessary in order to decide, which database would give sufficient data for an evaluation.

The following table shows the composition of the publication data from 2012/13 – 2015 with respect to the allotted research fields. The information of the researcher’s disciplines is taken

from the staff list of the official website of the research center. This information served as a basis to transform it into the subject categories of Web of Science.

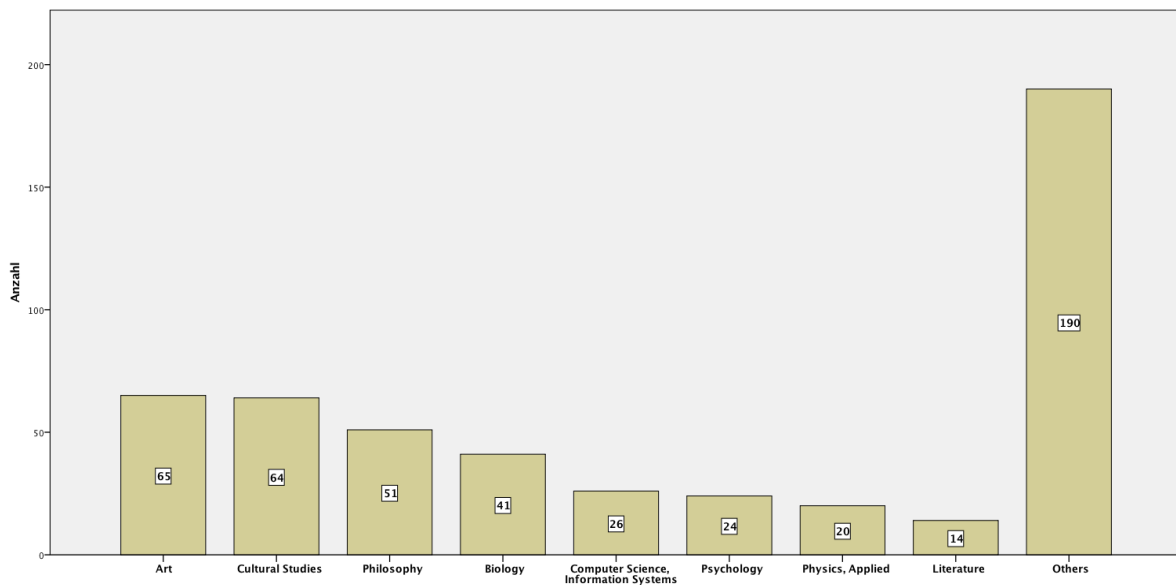


Figure 1: Composition of disciplines in the publication list of the research center BWG

This bar chart above (Figure 1) indicates, that most of the publications can be allotted to arts and humanities. Followed by disciplines from the sciences (Biology, Physics etc.) and also Psychology, which is counted as social science in the ISI citation index. There are a lot of other disciplines, but which much less frequencies. Therefore, they are summed up in the bar chart as “Other”.³

In order to have a more broad and comprehensive view on the composition of the BWG publications in relation to its disciplines, I transformed the values for the disciplines into the subject categories of Web of Science and allotted these categories to the three main research fields (science, social sciences and arts and humanities), based on ISI’s Citation Indices (Thomson Reuters 2016). In order to have explicit terms, I refer to these three categories as research fields in all calculations.

³ For a detailed view of the frequencies of the disciplines see attachment

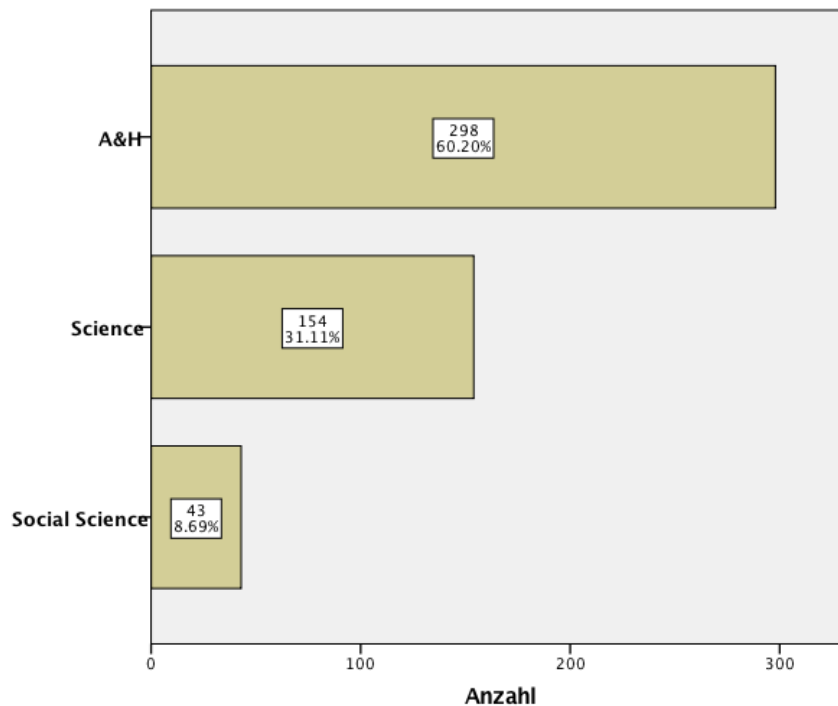


Figure 2: Proportion of research fields in the publication lists

This diagram (Figure 2) shows that most of the publications are from the arts and humanities with about 60%. Publications from the sciences make up about 31% of the publications list. Circa 8,7% of the publications are from the social sciences.

Another important aspect for the consideration of the right citation database, are the publication types of the publications.

According to Mongeon (2014), there is a shortage of journals in the fields of arts and humanities compared to social sciences and the life and natural sciences. Moreover, there is a more considerably lack of journals in German language. Also, monographs and edited volumes, which still are important publication formats in humanities and arts, are not extensively indexed by citations databases like Web of Science. As Münz (2012) states, in the Social Science Index of WoS there is still little coverage of German titles:

“Dabei ist das Grundproblem die Repräsentativität der bibliometrischen Daten. Der oft verwendete Social Science Citation Index erfasst weit überproportional englischsprachige Fachzeitschriftenaufsätze und damit nur einen Bruchteil der realen, in den Sozialwissenschaften noch stark national und in Sammelbänden und Monographien stattfindenden Diskurse.” (Münz 2007)

This Figure below (Figure 3), taken from „Scientific impact quantity and quality: Analysis of two sources of bibliographic data” (Belew 2005 p. 9), shows the difference of coverage of publication types between Google Scholar and ISI Web of Science.

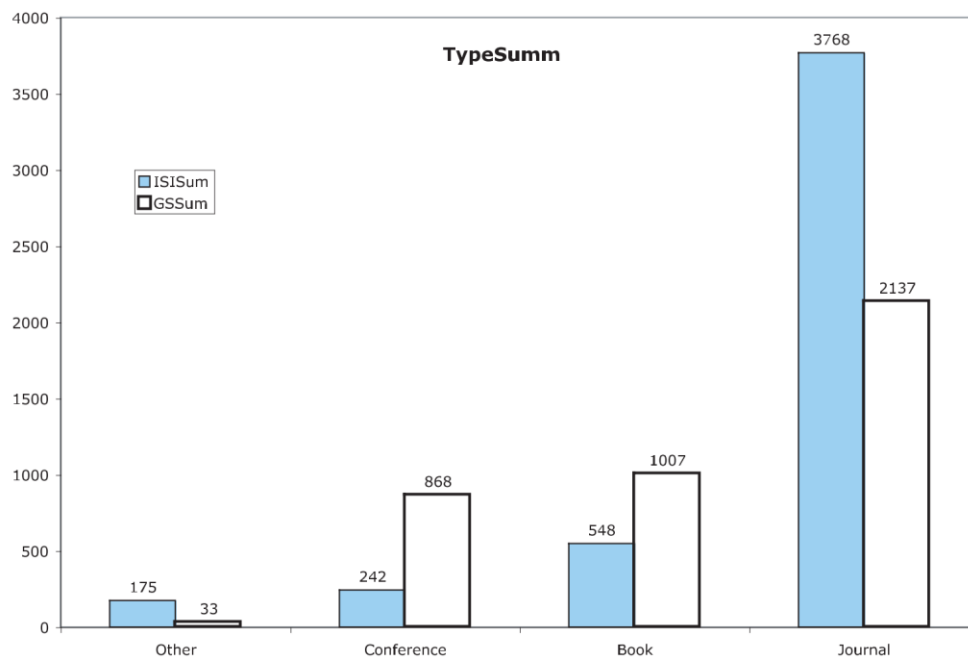


Figure 3: Publication types covered in GS and WoS in a coverage study by Belew, R.

In his paper, Belew (2005) criticizes the use of the ISI-based database in certain points. Not only, was there a lack of conference proceedings and books, but also regarding journals, he states, that ISI only covers the “most important” ones and not as many as possible (Belew 2005).

With respect to the types of publications, this following bar chart shows the format of research publications in three categories: journal, and edited collections. The bar chart below (Figure 4) shows the proportions of the publication formats in the publication lists:

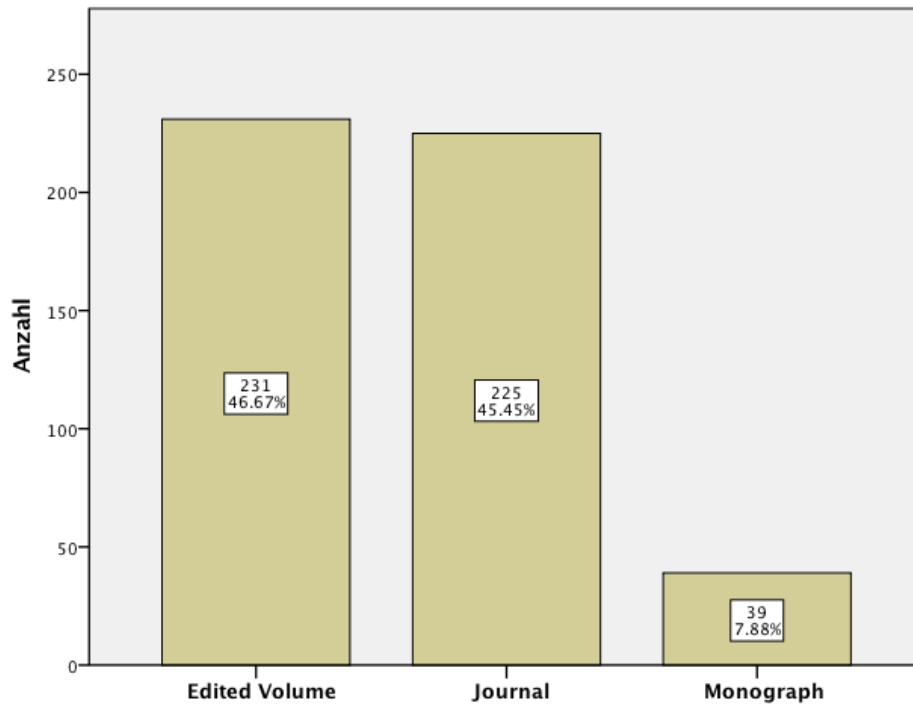


Figure 4: Proportion of publication formats in the publication lists

Since, edited volumes and monographs are both book publications, the majority of the publications are books and less than half of the publications are journal articles. Conference proceedings were also categorized as edited collections, because many authors contribute articles to a specific topic. Although one can argue that some contributions to a conference proceedings might as well be published as a paper in a separate journal, but that can also be true for book chapters. When comparing the numbers with the coverage study above between Google Scholar and ISI, one can refer that Google Scholar has the potential to provide more data, concerning the main publication format of all publication in the data set.

Another fact about the coverage is the time, which is needed for the database to index the latest publications, for this examination deals with new publications within the years of 2012/13 until 2015.

Furthermore, one principle stated in the publication policy of the research cluster is to publish open access. Considering this aspect, a data source should be selected which also covers open access publications.

As seen Google Scholar provides a lot of data. With its inclusion of Open Access-journals, it poses a potential candidate for a coverage study (Mayr 2009).

Finally, also Bornmann (2015) recommends not using Scopus or WoS for the research fields of humanities and the social sciences because these fields are often “nationally and regionally” oriented (Bornmann 2015). This would mean, that these publications are not sufficiently covered in WoS and Scopus.

This claim is also supported by Horst Bredekamp, the founder of the Cluster of Excellence Bild Wissen Gestaltung and prominent author in art history, which plays an important part in the interdisciplinary research work: “Weil seit 1830 ... die Kunstgeschichte im deutschen Sprachraum zuerst zum Universitätsfach geworden ist, hat von etwa 1830 an die deutsche Sprache eine gewisse Dominanz erlangt.” (Bredekamp 2012 p. 201).

Another argument Bornmann provides, is the already mentioned aspect of publication behavior in the humanities and social sciences:

“In the area of humanities and social sciences, publication tends predominantly to be in the form of books or monographs, which are essentially excluded as database documents (source items) for WoS or Scopus. Thus, typical publications in the humanities and social sciences are only insufficiently captured by these databases” (Bornmann 2015)

On the other side, Google Scholar is also known to index books (Bornmann et al. 2016). Its important project and service Google Books is one example of the coverage of books. Together with its comprehensive indexing of academic publishers Google Scholar can be seen as more advantageous than WoS or Scopus (Jacso 2008a)

These quotations prove that WoS and Scopus are not suitable for a bibliometric analysis with the given publication data of the research center BWG. So, it can be said, that Google Scholar has the best potential to sufficiently cover the publications.

With the given arguments above, it can be assumed, that Web of Science will not cover the publications of the research cluster as extensive as Google Scholar for the retrieval of bibliometric indicators. Also, Google Scholar can be a useful database for exploring information, which is not covered in the traditional database like Web of Science.

Because of the given arguments, I decided to get the necessary data from Google Scholar in order to measure the coverage and conduct a bibliometric analysis. The author name searches in Web of Science will be used to prove, that there is even less data, concerning the BWG publications in this traditional citation database. This is one reason, why I suggest neglecting bibliometric indicators in the research assessment of this Cluster.

As discussed in the previous chapter, a citation analysis with GS will be more informative, when enough time has passed. In this case I argued, that at least five years should pass before conducting an elaborate research assessment with bibliometric indicators and that it should be considered, that all publication must be at least two years old. Only then, the measurement of the impact of research works with quantitative can be more informative and comprehensive.

Nevertheless, in the next chapter I outline the approach and the method of the data analysis. I will describe all aspects of the data collection and the processing of the data with the variable in the calculation tables. Additionally, I will discuss the problems relating to the cleansing of the data.

4.2. Queries and Variables

This chapter presents the method and approach to the research question. Also, all findings of the bibliometric study will be presented and discussed. After having selected the databases, namely Google Scholar and Web of Science, the next step is to create the calculation table. All calculations, evaluations and visualizations have been done with Microsoft Excel (Microsoft 2014) and SPSS (IBM 2015). For reproducibility, the data sets used in this examination will be provided on external digital medium.

Another basis of this study is the staff list, which is publicly available via the official website of the research center (Bild Wissen Gestaltung 2016b). This data source is used to determine the disciplines of the authors. As mentioned, the information about the disciplines are again transformed into the three main subject categories of the ISI-database, namely science, social science and arts and humanities. The more detailed categorization of the disciplines from the official website has also been used for the calculation above (Figure 1). This has showed the proportions of the main disciplines of the given publications. In this case the result has been necessary to find the right citation database. Because some publications have a multidisciplinary character, so that more than one discipline could be assigned to the publications, I decided to transform the disciplines into broader research categories, eliding some interdisciplinary publications. Nevertheless, the detailed information about all publications regarding the disciplines, including multidisciplinary publications can be found in the attachment.

When transforming the discipline into the ISI Citation Index categories, I used only one category also for publications, which could not easily be allocated to only one category due to its

interdisciplinary character. This is not only necessary for identifying the proportion of disciplines of all given publications, but also for the purpose of calculating citation rates per research field, as I will show in an own chapter.

As said, the basic approach of retrieving bibliometric values in this study consists of searches in Google Scholar and Web of Science. This study can be divided into two sections, namely the coverage of academic titles in citation databases and the extraction of bibliometric indicators. Furthermore, the searches can also be differentiated between a title search and an author name search.

So far, the title query allows the creation of two new variables, namely the coverage variables for titles and the variable for citations.

In turn, the author name searches result in the values for the h -index and the information about the coverage of authors in the two databases GS and WoS.

As seen, there is not much need for cleaning the result list of the search engine, when querying publication titles, since the only values to set are “yes” or “no” for the coverage study. The level of this variable is either categorical or processed to numerical values for other calculations.

In the next step, I queried every name of the authors, who are member of the Cluster in both citation databases Google Scholar and Web of Science. Other coauthors, who are not affiliated with the research center, are respectively not subject of this study. The aim of this method is to retrieve the h -index from the two databases GS and WoS. This study takes more effort in both, the ensuring of the right query and the data cleansing of the result list of the search engine. However, the data cleansing part of this examination follows in the next chapter, with critical arguments and issues regarding the cleansing and the refinement of the result list of author-name-queries.

This means for the calculation, that two new variables can be created, one for the h -indices in Google Scholar and the other one for Web of Science. Every absence of authors in the database is marked as missing value. There are also cases, where you can find authors with one or a few titles, but still with an h -index of zero, ensuring to calculate the overall coverage of the h -index in both databases.

As seen, four new variables are created with two different queries-types in two different databases. I will also show, that for the coverage of publications, there is no need to query all titles

in another database, since with the coverage of *h*-indices in another database like Web of Science, the general coverage of publications can be approximately inferred.

Considering the retrieval of the *h*-index, all author names in the staff list were queried in the format: last name, name using the option “exact match”. In some cases the exact match option was not used to consider different spelling variants. For the calculation of the *h*-index, I used a browser plug-in called “Scholar H-index Calculator for Google Chrome ad Firefox™” (Ianni 2010). This browser plug-in calculates the *h*-index and other author metrics like the *g*-index automatically and instantly after each query, based on the entries in the result list.

Beforehand, an overview of the used variables in the calculation table is necessary.

In the beginning, I started to create variables in Excel, based on the publication list of the research center. For this I used the following variables:

1. Author

Although, I also created variables for every coauthor in the publication list, calculations for this examination focus the one author, who is member of the Cluster of Excellence, because in some cases there were authors, who are not affiliated with the research center. Author name queries were conducted in the two databases GS and WoS to calculate the *h*-indices. In cases where more than one author are member of the research cluster, both authors were queried in the database. This means a differentiation between first author and coauthor was not made, since this thesis examines the usefulness of the main bibliometric indicators, namely citation of the publications and the *h*-index of authors in the publication list.

- **Title**
- **Pub-year**

The publications range from the year 2012/2013 to 2015. Because the research center has started its work, in November 2012, there are only few publications allocated to this year. Although in the original publication list, from the years 2012 and 2013, contained more publications, I only counted publications, from principal investigators, which are topical related to the alignment of the interdisciplinary research work and had a certain influence on the creation of the Cluster of Excellence. This ensures also a more adequate measurement of the publication

output. The problems with the identification of research publications for the measurement of research performance, is the subject of the next chapter.

- **Edited volume**
- **Journal**
- **Monographs**

The categorization of publications is done separately, but can mostly be inferred directly from the bibliographic information. Usually, bibliometric analyses are done with journal articles. Because one argument against the usage of bibliometric indicators in research assessment concerning the research center BWG is the high amount of book publications in the humanities and social sciences, it is useful to split the publications into journal articles and books. In this case, I also divided the category book into monographs (usually by one author) and edited volume (here I categorized contributions to volumes and the volume itself in the same category). As mentioned, this is rather a formal distinction, which again has an informative character. In fact, volumes and monographs will be treated as books throughout the calculations.

- **Citation**

As discussed in previous chapters, the citation indicator was created with the values retrieved from Google Scholar. Self-evidently this variable is scaled metrically. It is noteworthy to say, that there is the distinction of publication with the value zero and cases when the publications were not indexed in Google Scholar. For the citation analysis, it is important to identify covered publication without citations, beside the overall coverage in Google Scholar.

- **H-index GS**
- **H-index WoS**

These values of the metric variables are retrieved from GS and WoS. Author name queries were conducted in the two databases. As described, the two indicators are analyzed and compared with each other to determine the usefulness of these indicators in research assessment.

- **In Google Scholar**

This variable is the result of title queries in Google Scholar. All titles in the publication lists from 2012/13 until 2015 were searched in Google Scholar. The variable has only two values: 1 = covered in Google Scholar

0 = not covered in Google Scholar

With this variable, the coverage of publications in Google Scholar can be measured. Due to the update frequencies of Google Scholar, the result can increase, when more time has passed (Falagas 20107, p 339). The result of this study stems from the May 2016.

- **Discipline**

- **Research Field**

In bibliometric analysis, it is common to calculate citation rates with the respective disciplines or research fields. For this purpose, the discipline of each publication on the list has been enhanced with the information about the author's discipline from the staff list of the research center. As explained, the information about the disciplines have been transformed into the main three research areas based on three citation indices of Thomson Reuter. These indices are called art and humanities-, social science-, and science citation index. All disciplines in the staff list were searched in the indices and then transformed into the respective superordinate concept. However, the discipline variable is also used for more detailed data evaluations for example to decide, which citation database to use, as explained in a previous chapter.

4.3 Data Cleansing

In this chapter, aspects regarding data cleansing will be discussed. There are problems regarding the title- and the author-name-queries in the databases. In particular, there are two aspects of concerns, namely the definition of academic publications and the identification of authors with common names.

The first problem in the data cleaning process occurred, when I searched the titles of the publications in Google Scholar. Since the original lists were not always clean of non-academic research work, like blog entries or interviews in newspaper, I deleted these works from the calculation table. Other irrelevant entries for a bibliometric analysis were for example: Websites,

conference posters or DVD's. Although, these "publications" do pose relevant research output also in the context of an evaluation, they are not suitable for citation analyses.

In this respect, a definition of academic publication is useful to determine, which titles to use in the queries. This aspect gains more importance, when we consider the fact that Google Scholar also covers grey literature and websites, which often are blogs (Courtault 2010), (Haddaway 15).

In order to distinguish authentic academic publication from other research output, I used the recommendations by the library of the University of Mannheim (Benz 2006). Of the nine criteria for determining research publications, the main aspects I checked were the usage of references and the publication format. This verification ensures, that only those publications are queried, which are academically relevant for a citation database.

Another aspect concerning the title searches in Google Scholar is the identification of the right publication in the result list. When there were numerous entries in the result list, I used the option "exact match" offered by the *h*-index calculator and by the advanced search option of GS. This method returns all matches based on the exact order of the queried string. This minimizes the possibility of wrong results. Then, I identified the publication by the title, the author name(s) and the publication format.

Even more important and problematic is the identification of authors, when querying author names in citation databases, especially with common names. As Thompson (2009) states this problem is wide spread in bibliometrics:

"Methodological issues, such as misspelled author names, homographs (ie, scientists with the same names in different disciplines), inconsistent use of author initials, or author name changes are potential problems." (Thompson 2009)

Although Web of Science offers a search with author identifiers, not every author is registered by respective identifiers services. Moreover, Google Scholar does not offer author id searches. Only when authors register in Google Scholar and claim their publications, authors can be identified. But of all the authors in the publication lists, there were only few authors who could be identified because they have registered in Google Scholar.

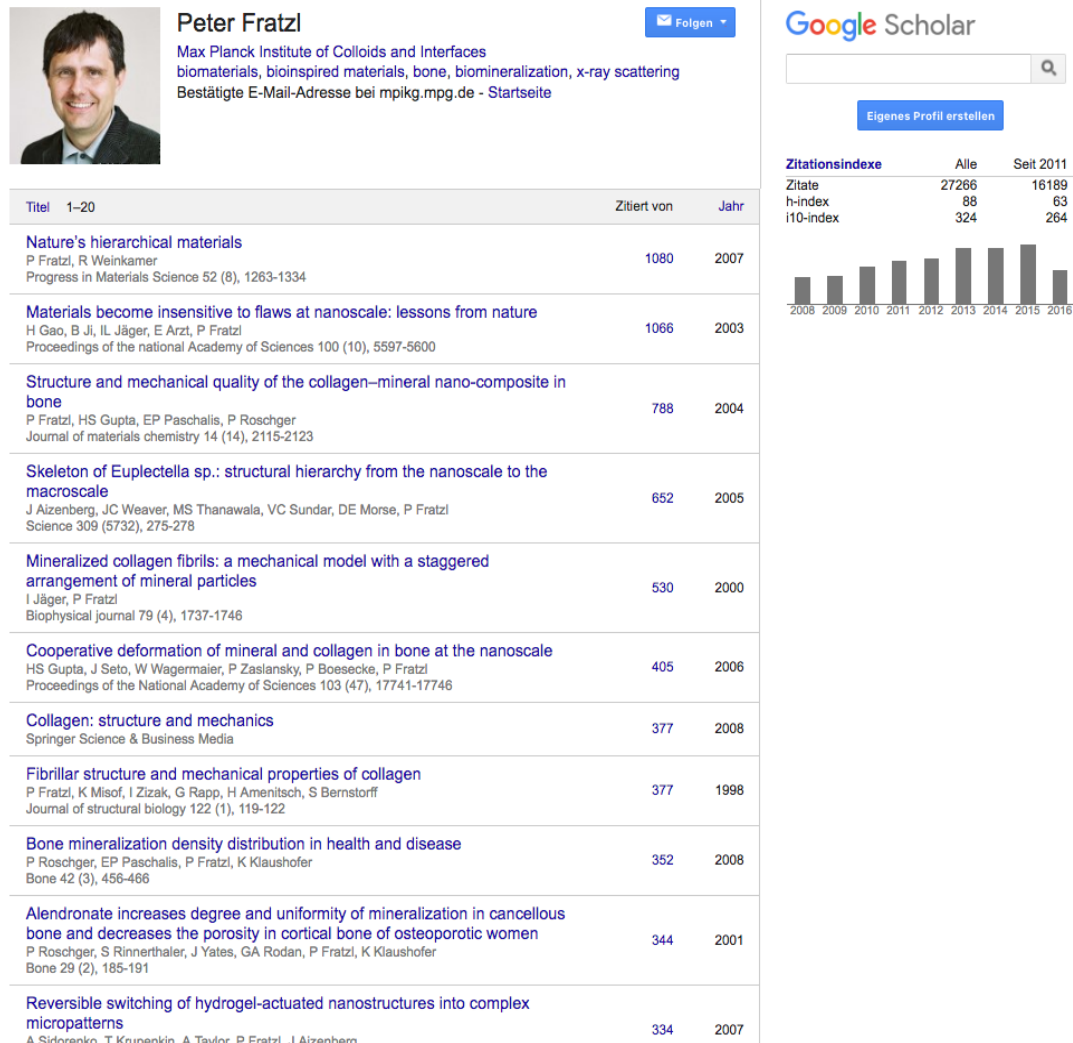


Figure 5: Profile of Peter Fratzl in Google Scholar (Screenshot)

As we can see in the screenshot (Figure 5), the numbers about citations are provided together with the h -index calculated by Google Scholar. But unfortunately, the registration of authors in GS are more exceptional than common praxis, so that bibliometricians and scientometricians often need additional software to calculate h -indices on the basis of GS data.

One advantage of retrieving the h -index in Web of Science is the option of author name searches with unique identifiers, avoiding any miscalculations of the h -index. WoS offers two author identifiers, namely ORCID (ORCID 2015) and ResearchID (Thomson Reuters 2016e), the latter is a service also owned by Thomson Reuters. Nevertheless, in practice the queries in WoS were done rather with the names, than with the identifiers, because, as I will show in the next chapter, only a small percentage of authors from the publication list are covered in Web of Science.

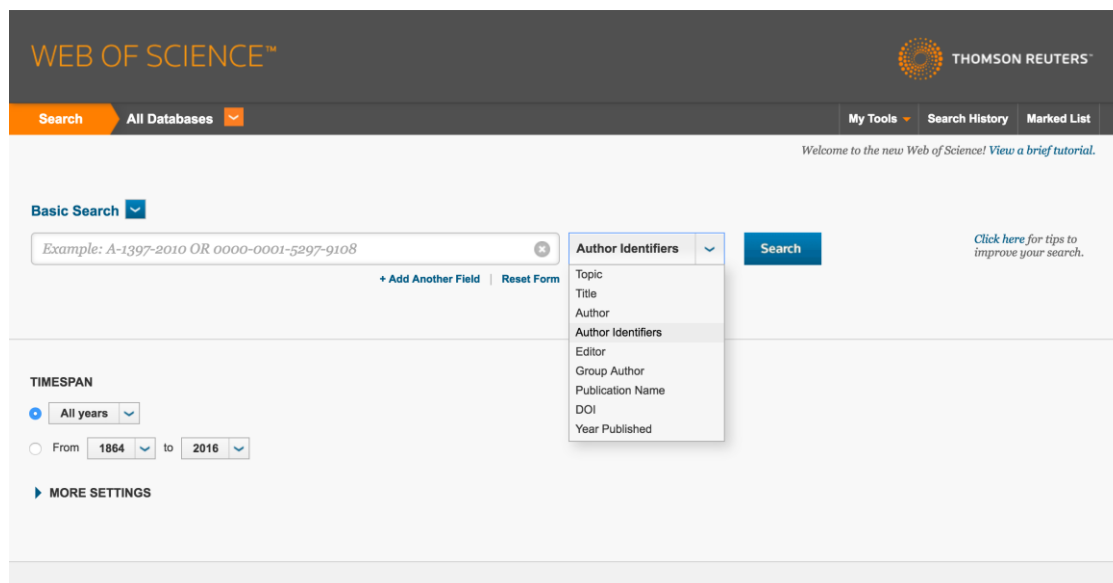


Figure 6: Author identifier search option in Web of Science (screenshot)

In this screenshot above (Figure 6), one can see the option to search with author identifiers in the drop-down menu of WoS. But for this examination, I have not used the option, since not all authors sign up for providers of identifiers. It would facilitate the author name searches, when author identifiers are more widespread in the academic community. But, as the following quote by Bornmann (2007) explains, not only must there be unique identifiers, but also other relevant information should be known to identify the author:

“The h index for a scientist can be found easily in Web of Science only if the scientist can be identified uniquely by name or if accurate publication lists can be pulled up in Web of Science by using a combination of the author name and address, or affiliation, search field”

(Bornmann and Daniel 2007)

On the other hand, results in Google Scholar for an author name search often contain irrelevant or false data, increasing the h -index sometimes to unrealistic heights, in particular when compared with the results from WoS. In these cases, often with common names, the identification of the author and his or her affiliated research work was the most important challenge in the cleansing process. The h -index calculator extension allows the use for an exact match search option. This option helped to identify the coverage of a specific publication and also the identification of the general works covered by a specific author, which is needed to calculate the h -index. Additionally, a thoroughly analysis of the result list was needed to exclude publications, which could not be allotted to the queried author. The calculator program offers the option to

de-select entries in the result list for this purpose. By excluding irrelevant and redundant entries in the result list, the calculator ignores these results and recalculates the h -index accordingly. The creator of the tool claims that the h -index is calculated based on the information displayed in the result list (Ianni 2010).

These cases only occurred with authors who have published a high amount of cited papers. New junior scientists, who also represent a big share in the publication list, where often not even covered in Google Scholar.

But for the other cases, I queried the author names in the ISI-Database, to compare the retrieval of the h -indices of the two databases.

Google Scholar search results for "author: "Macho, Thomas"".

Ungefähr 221 Ergebnisse (0,04 Sek.)

Redo the above query as: Quoted author name (selected) Word matching

Co-authors clusters

- J ASSMANN
- [57] GB Achenbach, ...
- [2] B Felderer
- T Macho
- [2] B Cornely
- T Macho
- [2] TH Macho
- A Truger
- [2] J Last
- T Macho
- [2] B Cole
- T Macho
- [4] W Braungart
- TH Macho, ...
- DS der
- [3] Doppelgnger
- E Duplication, ...
- [5] K Amann
- D Bogner, ...
- [2] T Macho
- M Wechsler
- [2] C Brndle
- TH Macho
- [4] E Habekost

Hide the calculator

Click here to remove this index formula

Normalization	Citations	h-index	g-index	e-index	delta-h	delta-g
none	1149	14	27	20	3	13
per co-authorship	909.1	13	23	16	3.5	25.8
per age	64.8	3	3	1	1.4	0.8

Data on video might not be enough for computing indices values.

Disable the advanced interface Reset selections Set Preferences to have Bibtex data and 20 results by default Help Page Donate

1. ☒ Keep 0 Invalid Citations 1 Authors

Todesmetaphern
 TH Macho - Zur Logik der Grenzerfahrung, Frankfurt/M, 1987 - tocs.ulb.tu-darmstadt.de
 Page 1. Thomas H. Macho Todesmetaphern 7,ur Logik der Grenzerfahrung Suhrkamp
 Page 2. Inhalt Einleitung: Vom Reden über den Tod 7 /. Konstruktionen der Todeserfahrung
 26 1. Natürlicher Tod 33 2. Freiwilliger Tod 47 3. Liebestod 62 II. ...
 Zitiert von: 137 Ähnliche Artikel Alle 2 Versionen Zitieren Speichern Mehr

2. ☒ Keep 0 Invalid Citations 2 Authors

[ZITATION] Weltrevolution der Seele
 P Sloterdijk, TH Macho - Ein Lese- und Arbeitsbuch der Gnosis von der ..., 1993
 Zitiert von: 59 Ähnliche Artikel Zitieren Speichern

3. ☒ Keep 0 Invalid Citations 1 Authors

[ZITATION] Von der Elite zur Prominenz. Zum Strukturwandel politischer Herrschaft
 TH Macho - Merkur 1993

Figure 7: Exact-match-option using the h -index calculator in Google Scholar (Screenshot)

The screenshot above (Figure 7) shows an author-name-search with the mentioned option of “quoted author name” by the calculator software. As mentioned, this search option can also be used in the advanced search option of GS. As one can see in the figure every entry can be unmatched as described above.

Additionally, the program offers an author list refinement tool to better identify authors, when there are more authors with the same name. Since this option only works, when the usual co-authors of the author are known, a deeper research of the authors works is necessary to identifier his or her publication colleagues. However, this option was of little use for this study, since a lot of authors from this research center are young scientists, who have not published many works with coauthors. To identify authors in GS, I used the exact match option and cleaned the result list or could make use of registered profiles in GS.

5. Bibliometric Analysis

This chapter involves all calculations and data evaluations with bibliometric measures regarding the publication data of the Cluster. First, the publication output in general will be analyzed. Besides the total amount of publications, these calculations incorporate publications per year, publications per author and publications per research field. When assessing the suitability of the citation-indicator and the *h*-index as bibliometric indicators, a publication analysis, without these indicators can be informative to evaluate the publication output in from different perspectives and gain more insights in order to better assess the mentioned bibliometric indicators.

In the next step, I provide information about the degree of coverage in Google Scholar. Also, I evaluate the coverage with other variables such as publication year, research field and authors. These evaluations will reveal details about the degree of coverage and lead to a broader understanding of the usefulness of the bibliometric indicators.

Further, a bibliometric analysis will be done with the retrieved data from Google Scholar. As previously mentioned, data from Web of Science is only used for a critical comparison of the respective data of the two databases, concerning the *h*-indices of authors. The analysis will focus on the citation data retrieved from Google Scholar. Although, it is debatable if a citation analysis with the amount of data retrieved from GS is appropriate, the citation variable will be calculated with the variables: publication year, author, and research field, in order to provide a more detailed view on the cited publication data.

Also, citation rates will be discussed in a separate calculation. The calculated citation rates are weighted by research field and publication year. The results will lead to the assessment of the citation indicator with GS data to measure the research performance of the Cluster.

Afterwards, the calculated *h*-indices from GS will be analyzed and compared with the data retrieved from WoS.

Finally, all results will be discussed, especially if these bibliometric indicators can be used in research assessment.

5.1. Publications

Before evaluating citation-based indicators, publication counts can also be seen as a general indicator of the research success of the Cluster of Excellence BWG. After having transformed the publication lists into one calculation table, I counted a total of 495 publications.

5.1.1. Publication per year

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2012	20	4.0	4.0	4.0
	2013	95	19.2	19.2	23.2
	2014	186	37.6	37.6	60.8
	2015	194	39.2	39.2	100.0
	Total	495	100.0	100.0	

Table 1: Publication output per year

As it can be seen in the table above (Table1) there are 495 publications in total. There is a steady increase of publication until 2015 with the most publication of 194 compared to 186 publications in 2014. Also, circa 39% of the publication are recent publications from the year 2015. In order words: most publications were published in 2015. This should be kept in mind, when calculating citation rates for each year and per research field.

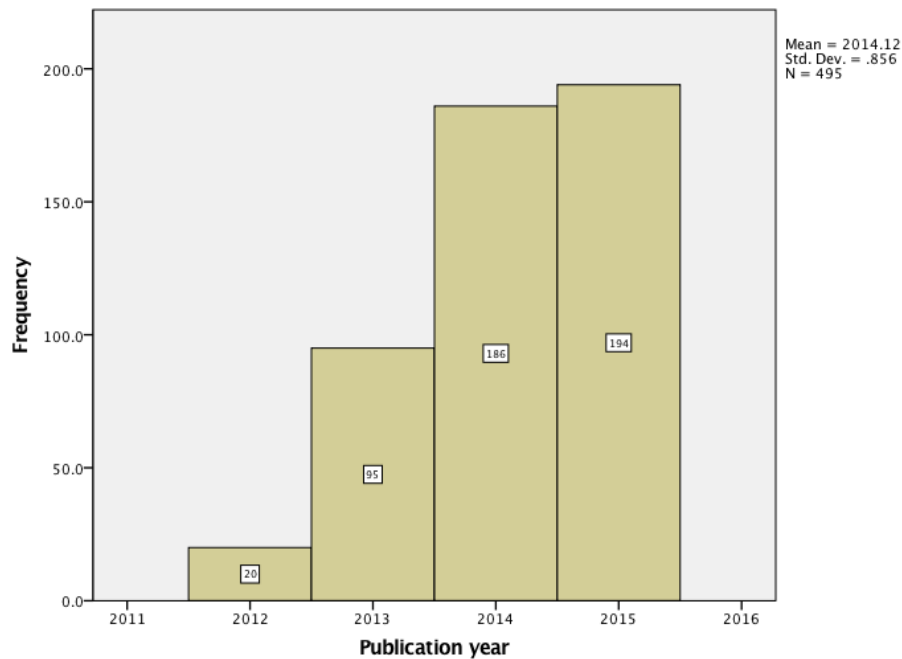


Figure 8: Publication output per year

In the bar chart above (Figure 8) the increase of publications per year is visualized. The jump of publication output in 2014 and 2015 compared to the beginning years in 2012/13 can clearly be seen. But already a certain slowness of increase can be recognized between the years 2014 with 186 publications and 2015 with 194 publications.

5.1.2. Publications per Publication Format

Because bibliometric studies are usually done with journals, I provide data about the distribution of the publication formats of the BWG publications.

	Publication Year				Total
	2012	2013	2014	2015	
Publication Format: Monograph	1	2	20	16	39
Journal	10	54	89	72	225
Edited Volume	9	39	77	106	231
Total	20	95	186	194	495

Table 2: Publication per year distinguished by type of publication

This table (Table 2) shows the numbers of publication per year distinguished by publication type. It can be seen, that of the 495 publications in total, only 225 publications have been published as journal articles. This must be taken into account, when attempting to collect bibliometric data from citation databases. As I discussed in chapter 3 and 4, citation databases preferably index journal articles. Because monographs and edited volumes are both book publications, there are 270 book publications compared to 225 journal articles. This means, that less than half of the publications are journal articles.

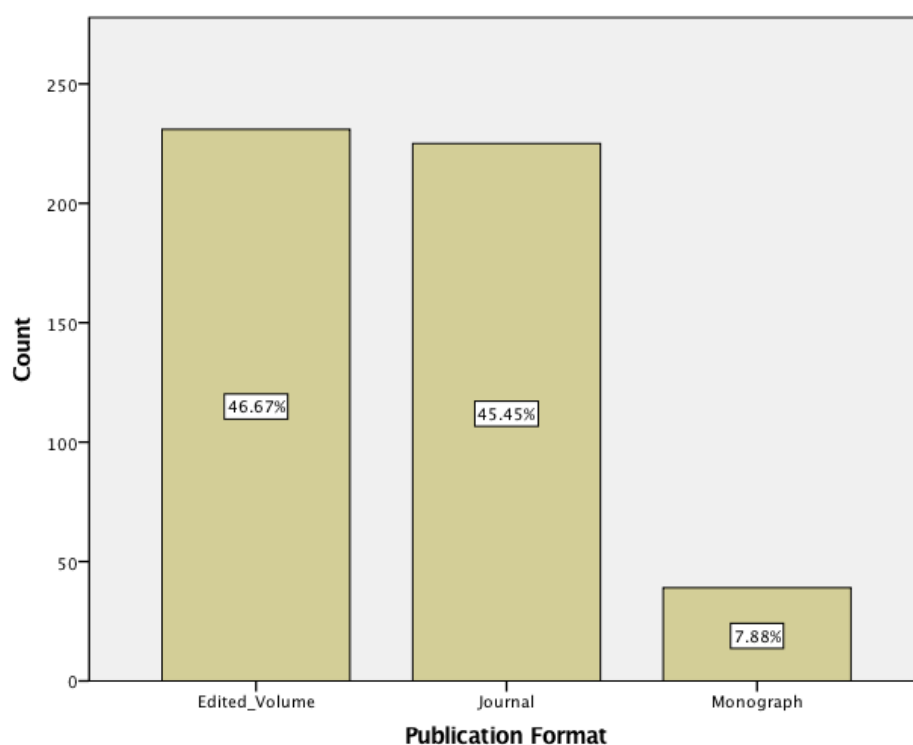


Figure 9: Proportions of publications formats in the data set

In the bar chart above (Figure 9), the proportions of publication format are visualized. Most of the publications in the data set are book publications, considering that both types edited volumes and monographs are published as books. Only about 45% of the publications from 2012/13 to 2015 are journal articles.

5.1.3. Publications per Research Field

Another aspect concerning the publication counts is the subject category. As described in chapter 4, I transformed the more detailed disciplines from the official website of the Cluster

into the three main research categories of the ISI-Citation Index. As the table below shows (Table 3), about 60% of the given publications belong to the arts and humanities.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	A&H	298	60.2	60.2	60.2
	Science	154	31.1	31.1	91.3
	Social Science	43	8.7	8.7	100.0
	Total	495	100.0	100.0	

Table 3: Publication per research field

As discussed previously in chapter 4.1, most publications derive from the arts and humanities with 298 publications in total, followed by publications from the sciences with 154 Publications (see Table 3). The social sciences are represented with 43 publications in the research output.

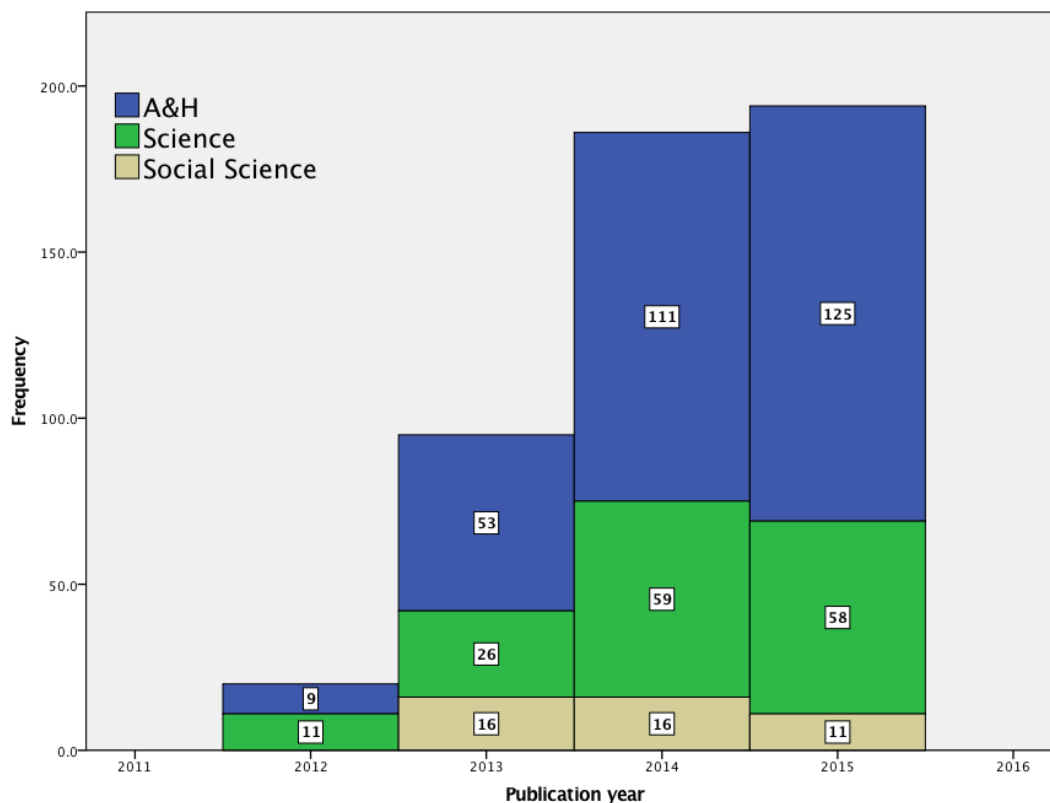


Figure 10: Publications per year splittet into the three main research categories

The bar chart in figure 10 shows the publications per year distinguished by the three research categories: science, social science and arts and humanities. There are only 20 publications counted for 2012, since the Cluster has been established in October 2012. Further, only those

publications have been counted for 2012, which contributed to the creation of the research center, for example the works from the arts and humanities, especially from the art history, as it is the subject of both speakers and founders of the Cluster.

It can also be seen, that the proportion of the three subject categories are very similar in the years 2013, 2014 and 2015.

5.2. Coverage

As described, a precondition for a bibliometric analysis is an adequate degree of coverage in citation databases. In the following the results from the coverage study are presented.

5.2.1. Total Coverage

The results of the total coverage of publications in GS as table and pie chart:

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid no	275	55.6	55.6	55.6
yes	220	44.4	44.4	100.0
Total	495	100.0	100.0	

Table 4 Coverage of the BWG publication data in Google Scholar

The table above (Table 4) provides descriptive statistics concerning the frequencies of the values “no” (=not covered in GS) and “yes” (=covered in GS) with absolute figures and proportions in percent. This counts for all calculations in this coverage study. It can be seen, that 44,4% of the 495 publications is covered in GS. This means 220 publications could be found in GS with respective citation counts. 275 publications or 55,6% are not covered in Google Scholar.

The pie chart below visualizes the coverage degree:

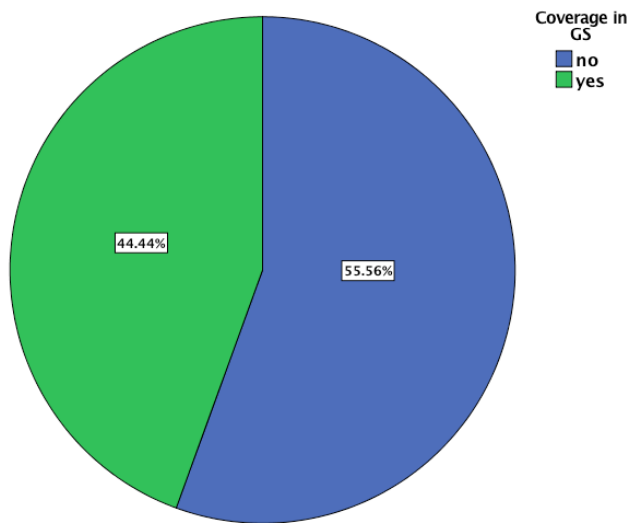


Figure 11: Pie chart visualizing the degree of coverage in Google Scholar

This visualization above (Table 11) shows the proportion of the covered publications and the publications, which are not covered. Less than half of all publications (ca.44%) are covered in GS.

5.2.2. Coverage per year

In the following, I present results from the calculation coverage per time, in this case publication year. This shows how the coverage rate has developed from 2012/13 to 2015.

	Publication Year				Total
	2012	2013	2014	2015	
Coverage in GS no	13	57	80	125	275
yes	7	38	106	69	220
Total	20	95	186	194	495

Table 5: Publication coverage in GS (total) per year

In this table above (Table 5), it can be seen, that there is an increase in covered publication in Google Scholar until the year 2014. From 2014 to 2015 there is a decrease in coverage from

106 to 69 covered publications. However, these numbers should be seen in relation to total publications output per year. The following bar chart shows the coverage rate per year.

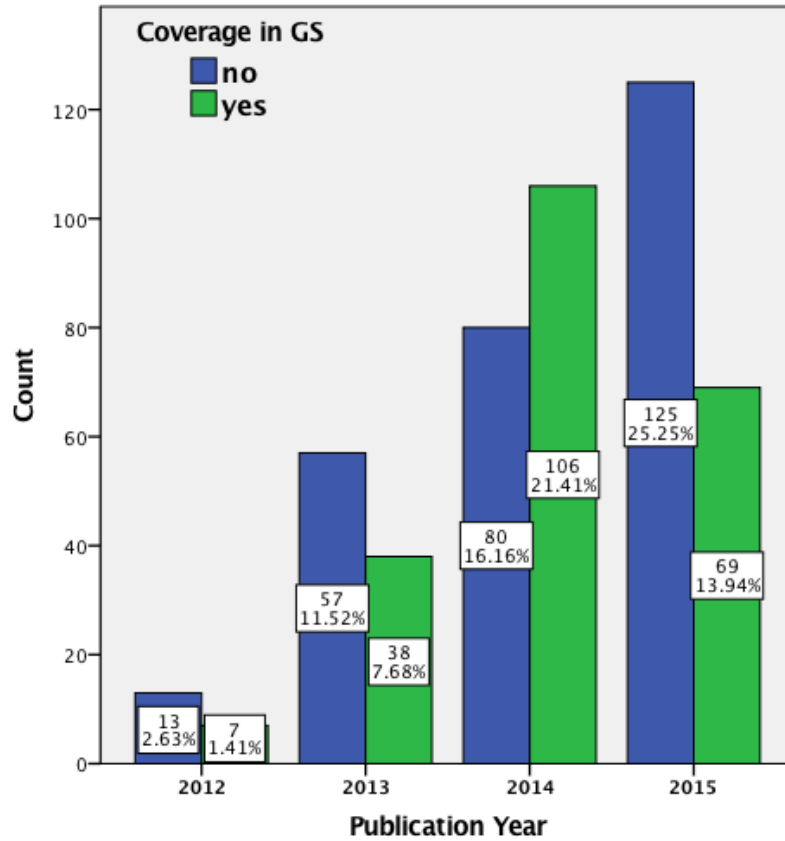


Figure 12: Publication covered in GS per year

In this bar chart above (Figure 12), the development of the coverage rate per time is visualized. It can be seen, that despite the increase of covered publications per year until 2014, there are less publication covered in 2013 than published for this year. The same counts for the year 2015. Only for the year 2014, there are more publications that can be found in GS, than the uncovered publications for this year.

5.2.3. Coverage of Research Field

As described, I transformed the information about the discipline to the three broad categories of ISI's citation indices: science, social science and arts and humanities. In the following, I provide information about the covered publication in GS distinguished by research field.

		Research Field			Total
		A&H	Science	Social Science	
Coverage in GS	no	192	67	16	275
	yes	106	87	27	220
Total		298	154	43	495

Table 6: Research fields of the BWG publications covered in Google Scholar

The table above shows that arts and humanities are the most covered publications in Google Scholar with 106 publications in total. From the 154 publications from the sciences, GS has indexed 87. Also, it can be seen, that from the 43 publications from the social sciences 27 of the publication can be found in GS. It can be said, that approximately one-third of the publications from the arts and humanities are covered in GS.

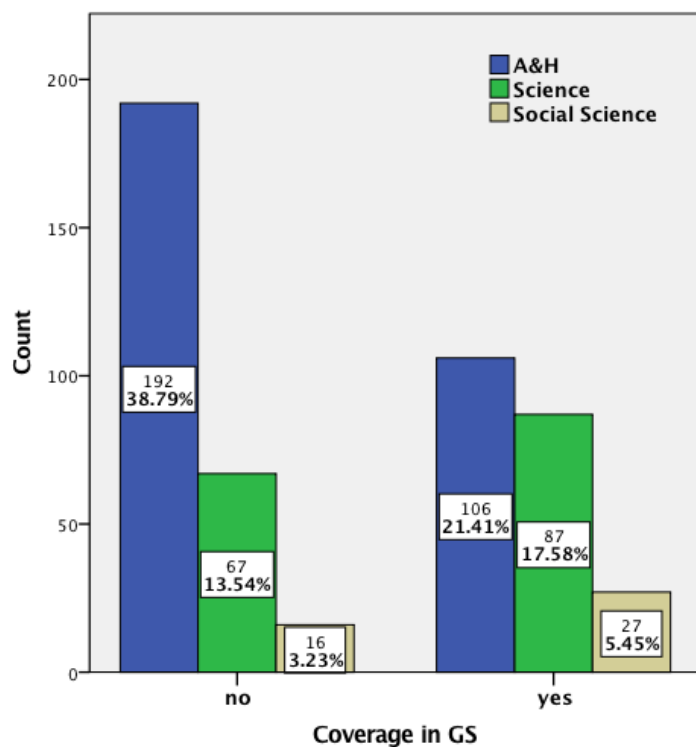


Figure 13: Coverage of the publications distinguished by the three research categories

The figure above (Figure 13) visualizes the degree of coverage concerning the Cluster-publications, distinguished by the research field arts and humanities, science and social science. It can clearly be seen, that there are more publications from the sciences covered in GS, than not covered publications from that field. Also, most publications from the social sciences are covered in GS. Only publications from the arts and humanities are proportionally not as broadly covered as the other research fields.

5.2.4. Coverage per Publication Format

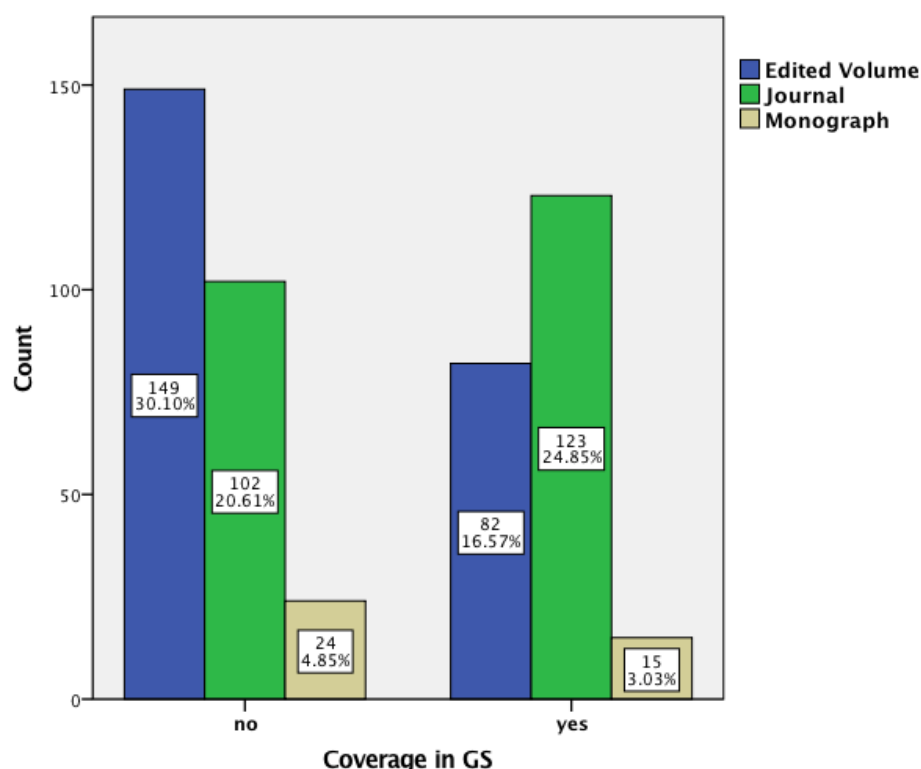


Figure 14: Coverage in GS per publication format

This figure above (Figure 14) visualizes the coverage of publications differentiated between the three main publication formats: edited volumes, journals and monographs. It must be taken into account, that both edited volumes and monographs are book publications. From the covered 44.4% publications, 24,85% are journals. 15,56% of the covered publications is book series or edited volumes. From the 25 monographs, about 3% are covered in GS. It can be seen, that only with the journal articles more publications are covered in GS, than not covered. Considering the other publication format less than half of the publications in those categories are covered in GS.

5.2.5. Coverage of Authors

In the case of the h -index in Google Scholar, I used (as mentioned in the respective chapter) an online h -index calculator. The following data provide information about the extent of indexed authors in the both databases. In order to find how many authors are represented in the publication list, I transformed the string-based author variable into a numeric variable, one figure for each author, starting with the value 1. Now we can calculate descriptive statistics and see the maximum, which is the number of authors in the publication list.

N	Valid	495
	Missing	0
Mean		58.79
Median		59.00
Mode		59
Range		111
Minimum		1
Maximum		112

Table 7: Descriptive statistics with the transformed author variable

As we can see in the table above (Table 7), the maximum of the values is 112. This means, there are 112 different authors in the publication list.

In order to calculate the covered authors in both citation databases, I decoded the respective two h -index variables into a binary variable, setting the values “no” = not covered and “yes” = covered for accordingly.⁴

⁴ see SPSS calculation tables on external medium “Data-CD”

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid no	398	80.4	80.4	80.4
yes	97	19.6	19.6	100.0
Total	495	100.0	100.0	

Table 8: Authors covered in Google Scholar

It can be seen in the table 8, that Google Scholar almost provides a total coverage of all authors of the publication list. From the 112 authors 97 authors can be found in GS.

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid no	439	88.7	88.7	88.7
yes	56	11.3	11.3	100.0
Total	495	100.0	100.0	

Table 9: Authors covered in Web of Science

In table 9 we can see that only 56 authors are covered in Web of Science. This is only approximately half of the authors. The coverage of authors in Web of Science can in turn be an indication of the total coverage of publication in Web of Science. As Google Scholar can be used to retrieve h-indices for almost all authors (ca. 90%) from the publication list, there is a total coverage of publications of 44,44%. On the other hand, Web of Science only covers about 50% of the authors of publications of the Cluster of Excellence BWG, inferring an even lower coverage degree of the BWG publications.

5.3. Citation Analysis

As described in chapter 5, there is an extent of coverage of the Cluster-publication to 44,44 % in Google Scholar. However, beside an evaluation with the variable research field, I also provide information about the total citation development per year and calculate an appropriate citation rate for the research fields, taking into account the limitations of the data.

5.3.1. Descriptive Data

In the following I present descriptive data on the citation variable:

N	Valid	495
	Missing	0
Mean		1.48
Median		.00
Std. Deviation		6.238
Range		65
Minimum		0
Maximum		65

Table 10: Descriptive statistics with the citation variable

In the table above (Table 10), it can be seen, that there is a maximum of 65. In fact, this high number can rather be seen as an outlier, since the arithmetic mean is 1.48 with a standard deviation of 6.232.

In the following, there is a boxplot of the citation variable for a better visualization of the distribution

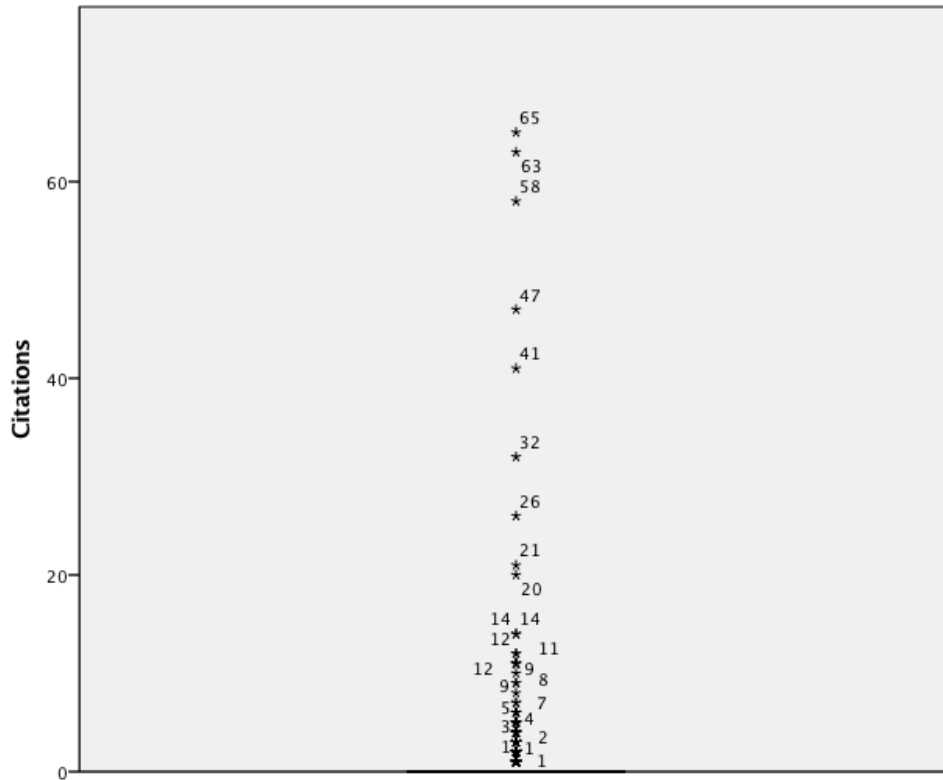


Figure 15: Boxplot of the citation variable

In Figure 15, we can see the distribution of the values of the citation variable. There are three values, higher than 50, namely 58, 63 and the maximum 65. However, as the arithmetic mean of 1.48 indicates, these high values can be seen as outliers. Before, calculating citation rates weighted by research field and publication year, I provide data of the citation values higher than zero. This can be useful, when assessing the applicability of the citation indicator in general.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	389	78.6	78.6	78.6
	> 0	106	21.4	21.4	100.0
	Total	495	100.0	100.0	

Table 11: Citation numbers higher than zero (> 0)

Table 11 shows the amount of citations with a value higher than zero. „0” stands for those publications with zero citations and when the publications could not be found in Google

Scholar. “> 0” stands for the publications with citations values higher than zero.

This means that there are 106 publications with a citation value 1 or higher.

In other words, from the 220 (44,44%) of covered publications in Google Scholar, there are 106 publications with a citation value higher than zero.

5.3.2. Citation Rates

One of the most important citation-based indicators in evaluative bibliometrics is the citation rate, which is a normalized value with factors such as publications per research field in a given time frame. Before presenting the results of the citation rates, I provide numbers about the citations per research field.

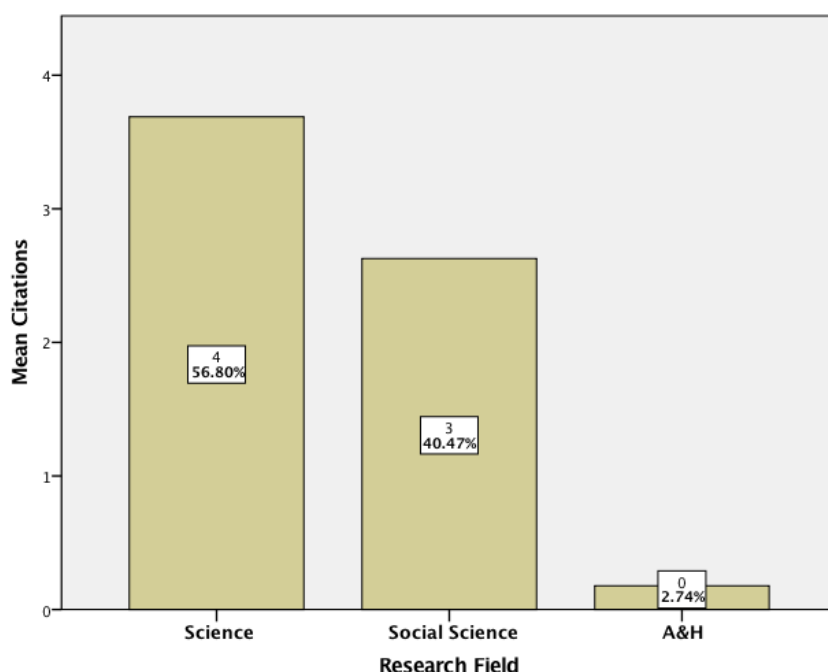


Figure 16: Citations (mean) per research field

It can be seen in the bar chart above (Figure 16), that science has the highest citations, followed by social science. The mean value for citations from the arts and humanities is 0 (below 1), although, this category makes up the biggest share of the publication list (about 3/5), as I have shown.

Before calculating citation rates for the research fields per year, I provide bar charts about the citations per year. The actual numbers, frequencies and distributions of the values are provided in the attachment.

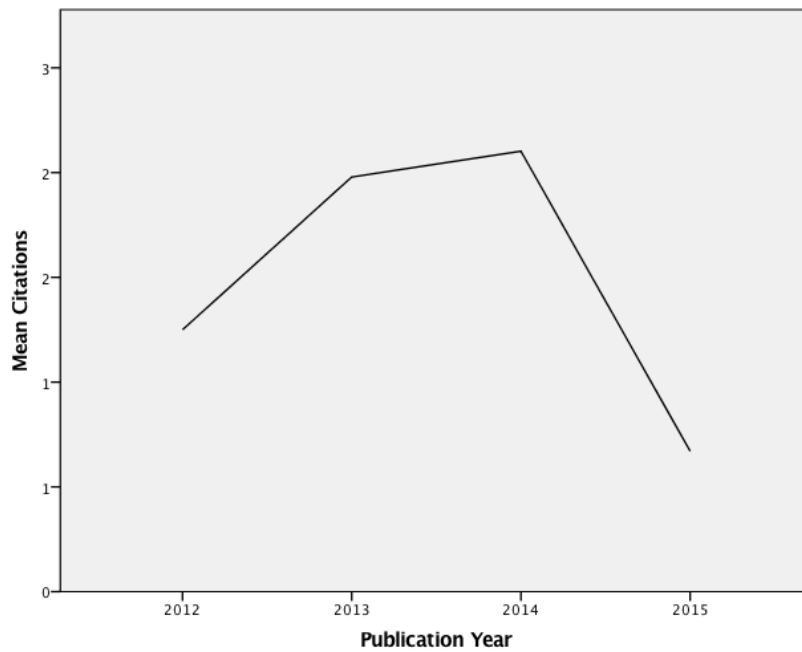


Figure 17: Citations (mean) per year

This graph (Figure 17) shows the citations per year. It also indicates the drop of citations from the year 2014 to 2015. Here the development of citations per year is measured with the arithmetic mean. Nevertheless, in order to calculate citation rates, the citation counts in total per research field is needed.

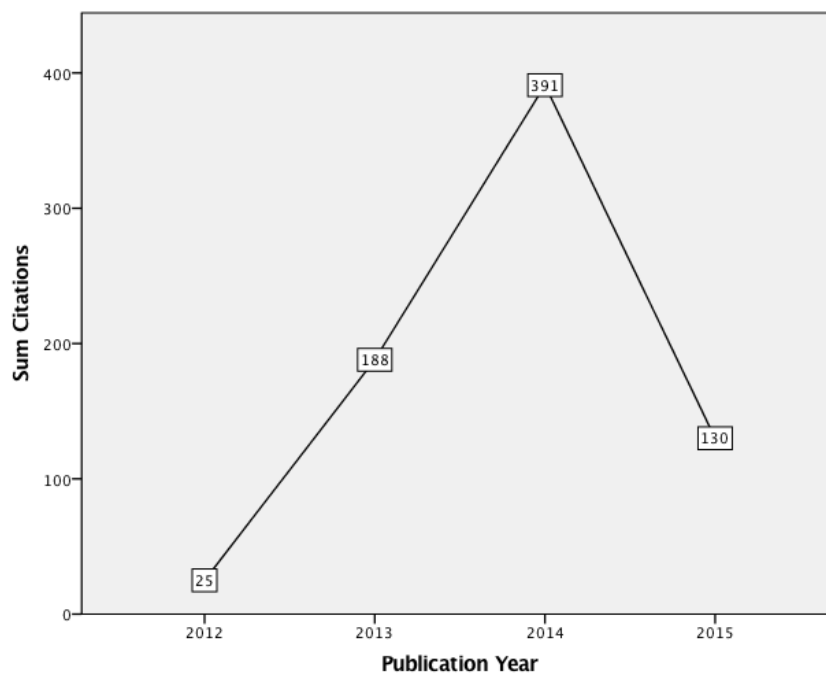


Figure 18: Citations in total per year

This charts above (Figure 18) show, that there is an increase in citation counts per year (in total) until the year 2014. Although, as I have showed, most publications were published in 2015 compared to the other years, there is a drop of citations for the year 2015, both in sum and for the mean.

To show an even more detailed aspect of the citation development, a normalization of the citation rate per research field is necessary. For this purpose, I divided the sum of citations per research field per year with the amount of publications from the respective research field per year.

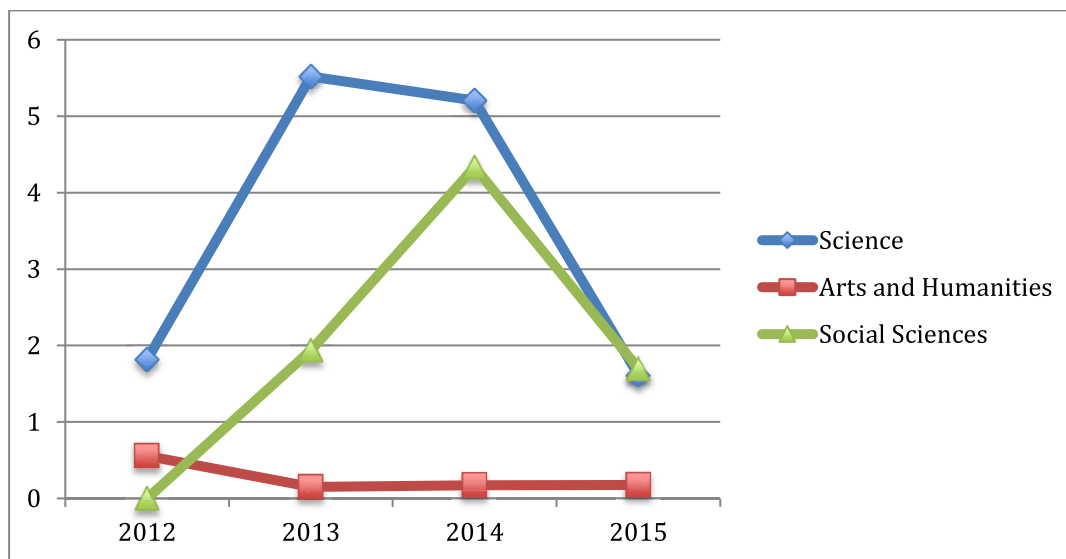


Figure 19: Citation rate curves for the three research field categories

It can be seen (Figure 19), that the citation rates for arts and humanities are relative low (0,5-0,17), compared to the other two research categories. Although, as I have shown, most publications are from the arts and humanities and in absolute figures there are also more publication from this category covered in GS, the citation rate is far lower, than for the other two research categories.

For the category science, the highest value for the citation rate can be found in the year 2013 with about 5,5. Then it decreases slightly in 2014 and falls to 1,6 in 2015.⁵ The citation rates for the research category social science increases until 2014 to 4,3 and falls to 1,7 in 2015.

⁵ All figures and numbers can be found in Excel calculation table on external medium (CD) attached to this thesis

5.4. H-index Analysis

As shown previously, from the 112 authors in the publication list, there are 97 found in Google Scholar. So, when we take into account the limitation of the citation-based indicator in this examination, due to the lack of coverage, the h -indices for authors calculated with Google Scholar could be instead used for a research evaluation, if one ignores the discussed weaknesses of this author metric. However, a glimpse at the h -index numbers in mean per research field category can give more insights of the usefulness of the data.

5.4.1. H-index GS

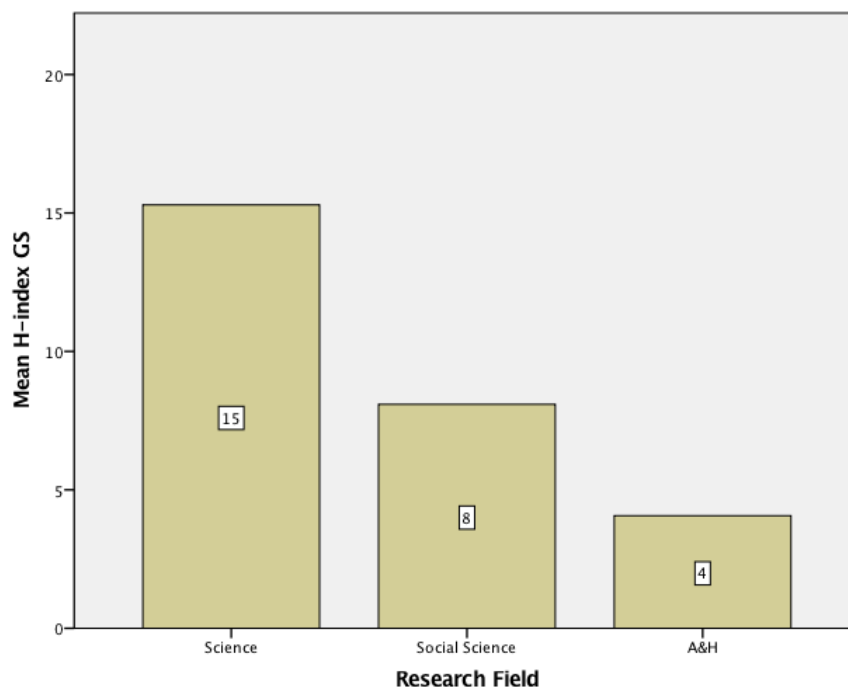


Figure 20: The mean value of the h -indices per research field for the data from GS

It can be recognized in the figure above (Figure 20), that for the h -index data calculated with Google Scholar, the numbers are high for the research field categories science followed by social science and the lowest value for the art and humanities. Again, the low h -indices for authors from the arts and humanities are contrasted by the dominance of this field in the proportion of research fields of the Cluster.

5.4.2. H-index WoS

Despite the shortage of h -index values from WoS, a distribution of the mean h -index per research field category can be helpful.

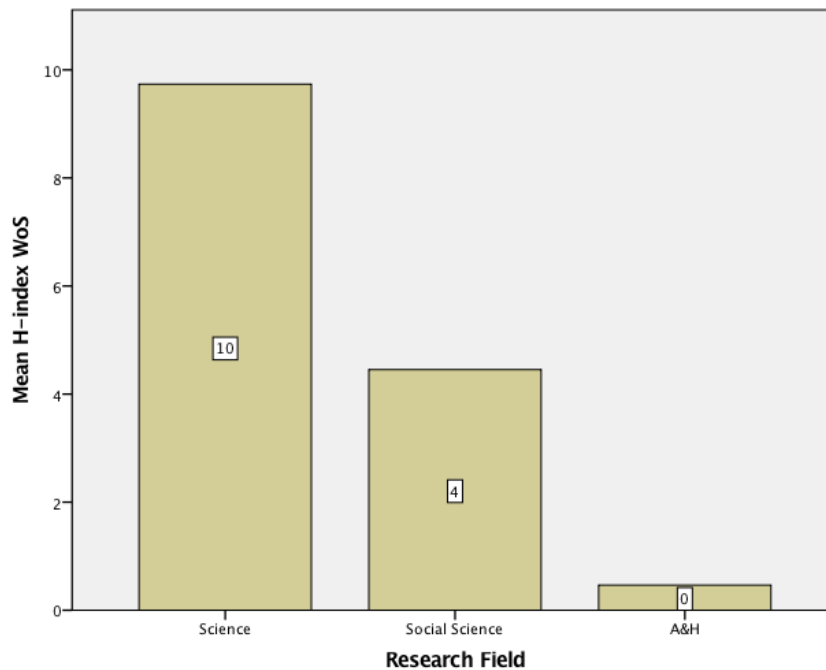


Figure 21: The mean value of the h -indices per research field for the data from WoS

Here (Figure 21), the h -index (mean) for the research field science and social sciences are very similar in terms of relation compared to the result with GS data. But as expected, there are far less German authors from the arts and humanities covered in GS, so that the h -index for this research field in WoS is 0.⁶

5.5. Discussion

As I have shown, there is a steady increase of publication output until 2015. The total amount of publications is 495. However, the numbers alone for the publications must be seen in relation to other values from other research center with a similar research focus. It is very unlikely to determine a similar research center, in particular when we consider, that the number must also

⁶ SPSS rounds up the figures in graphs. The actual number is therefore not important (0,17), considering the limitation of the data.

be normalized by the number of researchers and their research fields. Nevertheless, when we look at “Languages of Emotions” another Cluster of Excellence of the Freie Universität Berlin, there is the information about their total publication output from 2007 to 2014 on its official website (Freie Universität Berlin 2016a). The research cluster has an interdisciplinary composition with a focus on psychology, but also disciplines from the art and humanities are strongly represented. They also offer a database to search for the publications (Freie Universität Berlin 2016b). This is why the number given as total publication output for the years 2007 to 2014 (996) can serve as a point of reference. Whereas Bild Wissen Gestaltung has an output of 496 within 3 years and two months (2012/13 to 2015), the Cluster of Excellence Languages of Emotions has 996 publications within 7 years. If we also consider 7 years for the Cluster BWG, this research center has the potential to publish even more in the same time frame.

In fact, almost 40% of the publications were published in 2015. Therefore, especially for this year the publications should be covered in citation databases, but as I have shown the coverage of publications in GS for this year is even less than for the previous year. Only for the year 2014 there were more than half of the publications covered in GS.

Another aspect concerning the publication output is the format or type of the publications. Figure 9 shows that most of the publications are books. This is why the traditional ISI citation database WoS can be excluded as a data source for the publications of the Cluster BWG, because, as discussed, WoS does not cover books extensively enough, although the company has stated, that they are building a book citation index (Thomson Reuters 2016). But this index is not yet ready for bibliometric analyses (McKerlich 2013).

Also, relevant in bibliometric studies, is the discipline and the research fields of the publications. In Table 2, it can be seen, that about 60% of the publications are from the arts and humanities. This fact should also be considered, when deciding, which citation database to use as a data source for bibliometric indicators, because, as I have already discussed, bibliometric analyses are traditionally done with publications from the natural and life sciences (Moed 2009). Again, WoS does not seem to be the right database for a bibliometric analysis with the given data. Together with the mentioned argument, that most publications are books, also the citation database Scopus is not appropriate as a bibliometric data source. In their so-called Content Coverage Guide, the publisher Elsevier reports that most of the publications covered in its database are journals (Elsevier 2016b).

So, it can be said, that after evaluating the publications data, the only potential citation database for the retrieval of bibliometric indicators is Google Scholar, because of their coverage of books. Nevertheless, as I have shown, also GS only covers the publication of the Cluster to only 44,4%. Therefore, the coverage of the given publications in GS is not enough to be useful in research assessment. As I have calculated above, there is a difference between the coverage of publications of the years 2014 and 2015, although, GS claims that it adds documents “several times a week” (Google 2014). Only with existing data, the search engine needs up to 9 months to update the data. In this sense, the update frequency cannot really be a reason why the publications from 2015 are not as much covered as the publications of 2014, despite the fact that there are more publications in 2015. This means, only for citation numbers of existing records, the time factor plays an important role, as I have shown. This will be discussed later.

Also, there is a clear difference of the amount of covered publications per research field in GS (see Table 6). The dominant research category arts and humanities are approximately covered to one-third. Only more than half of the publications from the sciences are covered in GS. This means, that in general Google Scholar could be used as an alternative citation database only for the publications from the sciences. But because publication from the arts and humanities pose the biggest share in the publication list GS cannot be considered an appropriate database to retrieve the respective citation counts. This means in general, the total degree of coverage cannot be considered enough to measure the research performance or the impact of the whole research center.

I have shown, that most of the covered publications are journal articles. In this sense, more than half of the journal articles are covered in GS. On the other hand, less than half of edited volumes and monographs, the predominant publication format in the data set, are not covered in GS. This means that it is not recommendable to conduct bibliometric analyses, when most of the publications are books.

When it comes to the coverage of authors to calculate the *h*-index, Google Scholar indexes about 90 % of the authors. Therefore, GS can be used as a data source for the *h*-index indicator, despite the limitation of the coverage of the publication data. In fact, this limitation shows, that the *h*-index calculated with GS data should be used very carefully. Also, when we take into account the discussed weaknesses of this indicator in research assessment. In this context it must be emphasized, that it is not only “time consuming” to process the GS data, but also that

it could be “irresponsible” to only rely on data from GS for important research assessments (Jacso 2008c).

In chapter 5.2.5, I have presented the coverage of authors in WoS. Only about half of the authors could be found in WoS. This would mean, that even less BWG publications are covered in WoS. That is why, it can be concluded, that also WoS cannot be considered an appropriate data source for the publications of the research cluster.

In the citation analysis, it can be seen, that all publications from the dominant research field arts and humanities have a mean value of about zero (0,177). This can be explained, when calculating citation rates. As said, I have calculated the citation rate by dividing the citations per research field (in sum) with the amount of publications in that respective research field. Further, I calculated the citation rate for each research field per year. For the year 2015 the citation rates for science and social science drop from 5,2 and 4,3 to 1,6 and 1,7. Also the publication from the arts and humanities seem to stagnate around 0,17. One explanation for this development is that more recent publications need more time to reach the peak in citations. As we have already seen with this relative little data set with 495 records, the difference citation rate of the research categories can already be recognized.

I argued, that due the lack of coverage in citation databases, the bibliometric indicators should not be used in the research assessment of this Cluster. Also, having analyzed the citations, it can be concluded, that even if there was a broader coverage in citation databases, the citation indicators will be more useful, when the last publications are at least two years old.

6. Conclusion

In this Thesis, I examined the publication data of the Cluster of Excellence BWG with respect to creating bibliometric indicators. Therefore, this thesis contains a coverage study of the publication data in citation databases. As described in the respective chapters, I selected Google Scholar as the data source for the retrieval of citation counts and to calculate h-indices of the authors from the publication lists of 2012/13 to 2015. In order to collect bibliometric indicators, the publication data was queried in GS and WoS. The purpose was to analyze two important indicators in bibliometrics, namely citation rates and h-indices of the respective authors.

The examination was done from a critical standpoint towards an excessive use of quantitative indicators in research assessment. As discussed in chapter 3, there are several arguments against

such indicators in research assessment. Many researchers especially from the arts and humanities complain, that quantitative indicators shift the focus to performance and output rather than the qualitative aspects of research (Münz 2012; Jansen 2015; Kostoff 1998).

Furthermore, in bibliometrics there are different publication behaviors resulting in different citation frequencies. That's why elaborate formulas, calculations and normalizations must be conducted in order to compare the impact of the publications of the different disciplines with each other. Finally, citation-based analyses cannot be done immediately after the publications, since a certain time has to pass until the citation counts can be used. Depending on the research field and discipline more or less time must pass until citation rates reaches their peak. Concerning the given data of the research center BWG, the most recent data stems from the last year (2015) and because there is a significant share of arts and humanities, research assessment based on citation indicators would be too limited in their informative value.

Nevertheless, even with the publication data from 2013 and 2014, the coverage of publication in GS is as I have shown only ...%, with a total coverage of 35,1%. This means that there are other reasons for the unavailability of bibliometric data in citation databases. Aspects range from update frequencies of citation databases to publishing factors, such as research field or publication format.

I have shown that there is a significant part of books as a format of publication in the publication list of the research center BWG and that citation and bibliometric analyses are traditionally done with articles, which are more likely to get indexed in citation databases than books. Together with the regional context of publications, foremost the language of publications, the coverage of the BWG publications in GS is not sufficient for a bibliometric analysis. This can also be said with the coverage of authors from the publication list in the citation database Web of Science. This examination shows, that only 50% of the authors from the publication list are covered in WoS.

As I showed in Chapter 5.3, Google Scholar can be used to calculate the h -indices of the authors from the publication list. The author queries resulted in a coverage degree of h -indices of 90%. This means that Google Scholar could be used as the data source for the calculation of h -indices in the context of research assessment. Problems with data cleansing especially in the result list of GS must be taken seriously, as irrelevant results can distort the value of the h -index. Also, the problem with ambiguous author names must be addressed in order to get the valid value for the h -index. Unfortunately, GS does not support author identifier but offers a clear allocation

of publications to an author, with the option to register in Google Scholar and claim papers and publications.

As I have shown the citation rates for the three main research categories are very different, especially publications from the arts and humanities need more time to get more citations. Together with the degree of coverage of publications, bibliometric indicators should not be used for the research assessment of the Cluster BWG.

7. Bibliography

- Abbott, A. et al., 2010. Metrics: Do metrics matter? *Nature*, 465(7300), pp.860–862.
- Ball, R. & Tunger, D., 2005. *Bibliometrische Analysen - Daten, Fakten und Methoden: Grundwissen Bibliometrie für Wissenschaftler, Wissenschaftsmanager, Forschungseinrichtungen und Hochschulen*, Available at: <http://www.gbv.de/dms/bsz/toc/bsz116732865inh.pdf>.
- Barnett, G.A. et al., 1999. A mathematical model of academic citation age. *Journal of Composite Materials*, 33(10), pp.928–940.
- Belew, R.K., 2005. Scientific impact quantity and quality: Analysis of two sources of bibliographic data. *Cognitive Science*, p.12. Available at: <http://arxiv.org/abs/cs/0504036>.
- Benz, C., 2011. Kriterien wissenschaftlicher Publikationen. Available at: https://www.bib.uni-mannheim.de/fileadmin/elearning/informationskreislauf/kriterien_wissenschaftlicher_publikationen.html [Accessed July 3, 2016].
- Bild Wissen Gestaltung. Ein Interdisziplinäres Labor, 2016. Publikationen. Available at: <https://www.interdisciplinary-laboratory.hu-berlin.de/node/32> [Accessed July 3, 2016].
- Bild Wissen Gestaltung. Ein Interdisziplinäres Labor, 2016. Bild Wissen Gestaltung. Ein Interdisziplinäres Labor. Available at: Bild Wissen Gestaltung. Ein Interdisziplinäres Labor [Accessed July 3, 2016].
- Bornmann, L. et al, 2014. Bibliometric Standards for Evaluating Research Institutes in the Natural Sciences. In *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*. Cambridge: MIT Press, pp. 201–224.
- Bornmann, L., 2008. Scientific Peer Review : An Analysis of the Peer Review Process from the Perspective of Sociology of Science Theories. *Human Architecture: Journal of the Sociology of Self- Knowledge*, 6(2).
- Bornmann, L. & Daniel, H.-D., 2007. What do we know about the h index? *Journal of the American Society for Information Science and Technology*, 58(9), pp.1381–1385.
- Bornmann, L. et al., 2015. The application of bibliometrics to research evaluation in the humanities and social sciences : an exploratory study using normalized Google Scholar data for the publications of a research institute. *Figshare*, (February 2016).
- Bredenkamp, H., 2012. Deutsch in der Kunstgeschichte. In H. Oberreuter, ed. *Deutsch in der Wissenschaft - ein politischer und wissenschaftlicher Diskurs*. München: Olzog Verlag, pp. 201–205.
- Courtault, J.-M. et al., 2010. Research in economics and management in France: A bibliometric study using the h-index. *Journal of Socio-Economics*, 39(2), pp.329–337. Available at: <http://dx.doi.org/10.1016/j.socec.2009.12.006>.

- Deutsche Forschungsgemeinschaft, 2016. Excellence Initiative. Available at: http://www.dfg.de/en/research_funding/programmes/excellence_initiative/index.html [Accessed July 3, 2016].
- Deutsche Forschungsgemeinschaft, 2016. Cluster of Excellence. Available at: http://www.dfg.de/en/research_funding/programmes/excellence_initiative/clusters_excellence/index.html [Accessed July 3, 2016].
- Elsevier, 2016. Scopus. Available at: <https://www.elsevier.com/solutions/scopus> [Accessed July 3, 2016].
- Elsevier B.V., 2016. Scopus: Content Coverage Guide. , (January), p.28. Available at: https://www.elsevier.com/__data/assets/pdf_file/0007/69451/scopus_content_coverage_guide.pdf.
- Falagas, M.E. et al., 2007. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *The FASEB Journal*, 22(2), pp.338–342. Available at: <http://www.fasebj.org/cgi/doi/10.1096/fj.07-9492LSF>.
- Freie Universität Berlin, 2016. Languages of Emotions. Available at: <http://www.loe.fu-berlin.de/zentrum/publikationen/index.html> [Accessed July 3, 2016].
- Freie Universität Berlin, Languages of Emotion - Publikationen. Available at: <http://www.loe.fu-berlin.de/zentrum/publikationen/datenbank/index.html> [Accessed July 3, 2016].
- Gerhards, J., 2013. *Der deutsche Sonderweg in der Messung von Forschungsleistungen*, Berlin-Brandenburgische Akademie der Wissenschaften.
- Glänzel, W., 2006. On the h-index - A mathematical approach to a new measure of publication activity and citation impact. *Scientometrics*, 67(2), pp.315–321.
- Google Scholar, 2014. Search Tipps Coverage. Available at: <https://scholar.google.com/intl/en/scholar/help.html#coverage> [Accessed July 4, 2016].
- Haddaway, N.R. et al., 2015. The role of google scholar in evidence reviews and its applicability to grey literature searching. *PLoS ONE*, 10(9), pp.1–17.
- Havemann, F., 2009. *Einführung in die Bibliometrie*, Berlin: Gesellschaft für Wissenschaftsforschung. Available at: <http://nbn-resolving.de/urn:nbn:de:101:1-2009041608>.
- Hirsch, J.E., 2005. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci U S A*, 102(46), pp.16569–16572. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16275915>.
- Humboldt-Universität zu Berlin, 2016. Exzellenzcluster. Available at: <https://www.exzellenz.hu-berlin.de/de/exzellenzcluster-2> [Accessed July 3, 2016].

- Ianni, G. et al., 2010. Scholar H-Index Calculator for Google Chrome and Firefox.
- Jacso, P., 2008. Testing the Calculation of a Realistic h-index in Google Scholar, Scopus, and Web of Science for F. W. Lancaster. *Library Trends*, 56(4), pp.784–815. Available at: http://muse.jhu.edu/journals/library_trends/v056/56.4.jacso.html
<http://muse.jhu.edu/journals/lib-summary/v056/56.4.jacso.html>
http://muse.jhu.edu/journals/library_trends/v056/56.4.jacso.pdf.
- Jacso, P., 2008. The plausibility of computing the h-index of scholarly productivity and impact using reference-enhanced databases. *Online Information Review*, 32(2), pp.266–283.
- Jacso, P., 2008. The pros and cons of computing the h- index using Google Scholar. *Online Information Review*, 32(3), pp.437–452. Available at: <http://www.emeraldinsight.com/doi/abs/10.1108/14684520810889718>.
- Jamali, H.R. & Nabavi, M., 2015. Open access and sources of full-text articles in Google Scholar in different subject fields. *Scientometrics*, 105(3), pp.1635–1651. Available at: ["http://dx.doi.org/10.1007/s11192-015-1642-2](http://dx.doi.org/10.1007/s11192-015-1642-2).
- Jansen, D., 2015. Drittmittel als Performanzindikator der Wissenschaftlichen Forschung. *Drittmittel als Performanzindikator der wissenschaftlichen Forschung*, 1, pp.125–149.
- Khabsa, M. & Giles, C.L., 2014. The Number of Scholarly Documents on the Public Web. *PLoS ONE*, 9(5), p.e93949. Available at: <http://dx.plos.org/10.1371/journal.pone.0093949>.
- Kostoff, R.N., 1998. The use and misuse of citation analysis in research evaluation. *Scientometrics*, 43(1), pp.27–43. Available at: <http://link.springer.com/10.1007/BF02458392>.
- Kousha, K. & Thelwall, M., 2007. Google scholar citations and google Web/URL citations: A multi-discipline exploratory analysis. *Journal of the American Society for Information Science and Technology*, 58, pp.1055–1065. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/asi.20584/full>.
- Marx, W., 2015. Bibliometrie in der Forschungsbewertung: Aussagekraft und Grenzen. *PhD Proposal*, 1.
- Marx, W., Schier, H. & Andersen, O.K., 2006. Using time dependent citation rates (sales curves) for comparing scientific impacts. , p.9. Available at: <http://arxiv.org/abs/physics/0611284>.
- Mayr, P., 2009. Google scholar als akademische suchmaschine. *VOEB-Mitteilungen*, 62(2), pp.18–28. Available at: <http://www.ib.hu-berlin.de/~mayr/arbeiten/mayr-voeb209.pdf>.
- McKerlich, R., Ives, C. & McGreal, R., 2013. Measuring use and creation of open educational resources in higher education. *International Review of Research in Open and Distance Learning*, 14(4), pp.90–103.

- McVeigh, M.E., 2004. Open Access journals in the ISI citation databases: analysis of impact factors and citation patterns: a citation study from Thomson Scientific. *Thomson Corporation*, (October), p.25. Available at: <http://biblioteca.uned.es/lenya/bibliuned/search-authoring/docpdf/oacitations2.pdf>.
- Moed, H.F., 2009. New developments in the use of citation analysis in research evaluation. *Archivum immunologiae et therapiae experimentalis*, 57(1), pp.13–8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19219533>.
- Mongeon, P. & Paul-hus, A., 2014. The journal coverage of bibliometric databases: A comparison of Scopus and Web of Science. , pp.1–6.
- Münch, R., 2007. *Die akademische Elite. Zur sozialen Konstruktion wissenschaftlicher Exzellenz*, Frankfurt am Main: Suhrkamp Verlag. Available at: <http://www.amazon.de/exec/obidos/redirect?tag=citeulike01-21&path=ASIN/3518125109>.
- Norris, M., Oppenheim, C. & Rowland, F., 2008. Finding open access articles using Google, Google Scholar, OAIster and OpenDOAR. *Online Information Review*, 32(6), pp.709–715.
- Ochsner, M., Hug, S.E. & Daniel, H.-D., 2014. Setting the stage for the assessment of research quality in the humanities. Consolidating the results of four empirical studies. *Zeitschrift für Erziehungswissenschaft*, 17(S6), pp.111–132. Available at: <http://link.springer.com/10.1007/s11618-014-0576-4>.
- ORCID, 2015. ORCID. Available at: <http://orcid.org/> [Accessed July 3, 2016].
- Orduña-Malea, E. et al., 2014. About the size of Google Scholar: playing the numbers. *Computer Science > Digital Libraries*, (Fri, 5 Sep 2014), pp.1–42.
- Reuters, T., 2016. Web of Science. Available at: https://apps.webofknowledge.com/UA_GeneralSearch_input.do?product=UA&search_mode=GeneralSearch&SID=W1n1cWIRIFqjprraCI7&preferencesSaved= [Accessed July 3, 2016].
- Stellenbosch University, 2015. Bibliometrics and citation analysis. Available at: <http://libguides.sun.ac.za/h-index> [Accessed July 3, 2016].
- Thompson, D.F., Callen, E.C. & Nahata, M.C., 2009. New indices in scholarship assessment. *American journal of pharmaceutical education*, 73(6), p.111. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2769533&tool=pmcentrez&rendertype=abstract>.
- Thomson Reuters, 2016. Science Citation Index. Available at: http://ip-science.thomsonreuters.com/mjl/scope/scope_sci/ [Accessed July 3, 2016].
- Thomson Reuters, 2016. Citation Rates. Available at: <http://ipscience-help.thomsonreuters.com/incitesLiveESI/ESIGroup/fieldBaselines/citationRatesBaselines.html> [Accessed July 3, 2016].

Thomson Reuters, 2016. Book Citation Index. Available at: http://wokinfo.com/products_tools/multidisciplinary/bookcitationindex/ [Accessed July 3, 2016].

Thomson Reuters, 2016. Eugene Garfield: Sixty years of innovation in Information Science. Available at <http://thomsonreuters.com/en/articles/2015/eugene-garfield-sixty-years-information-science-innovation.html> [Accessed July 3, 2016].

Thomson Reuters, 2011. ResearcherID. Available at: <http://www.researcherid.com/Home.action?SID=W1n1cWIRIFqjprraCI7&returnCode=ROUTER.Success&SrcApp=CR&Init=Yes> [Accessed July 3, 2007].

Waltman, L. & van Eck, N.J., 2012. The inconsistency of the h-index. *Journal of the American Society for Information Science and Technology*, 63(2), pp.406–415.

Wang, J., 2013. Citation time window choice for research impact evaluation. *Scientometrics*, 94(3), pp.851–872.

Wissenschaftsrat, 2011. Empfehlungen zur Bewertung und Steuerung von Forschungsleistung. , p.49 S.

8. Attachment

1. Frequencies of disciplines based on the ISI subject categories.

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Acoustics	2	.4	.4	.4
Acoustics; Cultural Studies	10	2.0	2.0	2.4
Archeology; Art	3	.6	.6	3.0
Architecture	12	2.4	2.4	5.5
Architecture / Art	1	.2	.2	5.7
Architecture / Cultural Studies	1	.2	.2	5.9
Architecture / Urban Studies	1	.2	.2	6.1
Architecture; Art	1	.2	.2	6.7
Architecture; Art	2	.4	.4	6.5
Art	65	13.1	13.1	19.8
Art / Biology	1	.2	.2	20.0
Art / Urban Studies	1	.2	.2	20.2
Art History	3	.6	.6	20.8
Art; Literature	3	.6	.6	21.4
Art; Philosophy	6	1.2	1.2	22.6
Biology	41	8.3	8.3	30.9
Biology / Art	1	.2	.2	31.1
Biology / Materials Science, Composites; Physics, Applied	1	.2	.2	31.3
Biology; Cultural Studies	4	.8	.8	32.1
Biology; Cultural Studies / Philosophy	1	.2	.2	32.3
Biology; Materials Science, Composites	1	.2	.2	32.5

Biology; Medicine, General & Internal / Materials Science, Composites / Phys- ics, Applied	1	.2	.2	32.7
Chemistry, Applied	1	.2	.2	32.9
Communication	7	1.4	1.4	34.3
Communication / Biology	1	.2	.2	34.5
Computer Science, Information Systems	26	5.3	5.3	39.8
Computer Science, Information Systems / Cultural Studies	1	.2	.2	40.0
Computer Science, Information Systems / Material Science, Composites	1	.2	.2	40.2
Computer Science, Information Systems; Linguistics	8	1.6	1.6	41.8
Cultural Studies	64	12.9	12.9	54.7
Cultural Studies / Medicine, General & Internal; Physics Ap- plied	1	.2	.2	54.9
Cultural Studies; Communication	1	.2	.2	55.2
Cultural Studies; Literature	3	.6	.6	55.8
Cultural Studies; Philosophy	1	.2	.2	56.0
Cultural Studies; Philosophy / History	5	1.0	1.0	57.0
Ethnic Studies	5	1.0	1.0	58.0
Ethnic Studies; Area Studies	4	.8	.8	58.8
Ethnic Studies; Cultural Studies	7	1.4	1.4	60.2

Ethnic Studies; History	3	.6	.6	60.8
History	6	1.2	1.2	62.0
History / Cultural Studies	3	.6	.6	62.6
History & Philosophy of Science	3	.6	.6	63.2
History; Art	4	.8	.8	64.0
History; Cultural Studies	1	.2	.2	64.2
Information Science & Library Science	10	2.0	2.0	66.3
Literature	14	2.8	2.8	69.1
Literature / History & Philosophy of Sci- ence	1	.2	.2	69.3
Literature; Cultural Studies	2	.4	.4	69.7
Literature; History & Philosophy of Sci- ence	1	.2	.2	69.9
Materials Science, Composites	6	1.2	1.2	71.1
Materials Science, Composites / Biol- ogy; Medicine, Gen- eral & Internal / Physics, Applied	1	.2	.2	71.3
Materials Science, Composites / Phys- ics, Applied	3	.6	.6	71.9
Mathematics	11	2.2	2.2	74.1
Medicine, General & Internal	11	2.2	2.2	76.4
Medicine, General & Internal / Cultural Studies	2	.4	.4	76.8

Medicine, General & Internal; Cultural Studies	1	.2	.2	77.0
Philosophy	51	10.3	10.3	87.3
Philosophy / Mathematics	1	.2	.2	87.5
Philosophy; Language & Linguistics; Literature,	1	.2	.2	87.7
Physics, Applied	20	4.0	4.0	91.7
Physics, Applied / Acoustics; Cultural Studies	1	.2	.2	91.9
Psychology	24	4.8	4.8	96.8
Psychology / Biology	1	.2	.2	97.0
Sound Studies	4	.8	.8	97.8
Urban Studies	4	.8	.8	98.6
Urban Studies; Architecture	7	1.4	1.4	100.0
Total	495	100.0	100.0	

2. Frequencies of the citation indicator

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 0	394	79.6	79.6	79.6
1	17	3.4	3.4	83.0
2	27	5.5	5.5	88.5
3	13	2.6	2.6	91.1
4	9	1.8	1.8	92.9
5	8	1.6	1.6	94.5
6	4	.8	.8	95.4
7	2	.4	.4	95.8
8	1	.2	.2	96.0
9	2	.4	.4	96.4
10	1	.2	.2	96.6
11	4	.8	.8	97.4
12	2	.4	.4	97.8
14	2	.4	.4	98.2
20	1	.2	.2	98.4
21	1	.2	.2	98.6
26	1	.2	.2	98.8
32	1	.2	.2	99.0
41	1	.2	.2	99.2
47	1	.2	.2	99.4
58	1	.2	.2	99.6
63	1	.2	.2	99.8
65	1	.2	.2	100.0
Total	495	100.0	100.0	

3. Descriptive statistics of the h-index variable calculated with GS Data

N	Valid	109
	Missing	386
Mean		8.38
Median		3.00
Std. Deviation		12.771
Range		88
Minimum		0
Maximum		88

Hindex_GS

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 0	13	2.6	11.9	11.9
1	19	3.8	17.4	29.4
2	17	3.4	15.6	45.0
3	9	1.8	8.3	53.2
4	6	1.2	5.5	58.7
5	2	.4	1.8	60.6
6	5	1.0	4.6	65.1
7	2	.4	1.8	67.0
8	4	.8	3.7	70.6
9	1	.2	.9	71.6
11	1	.2	.9	72.5

12	1	.2	.9	73.4
13	3	.6	2.8	76.1
14	1	.2	.9	77.1
15	2	.4	1.8	78.9
16	4	.8	3.7	82.6
17	1	.2	.9	83.5
18	2	.4	1.8	85.3
19	4	.8	3.7	89.0
20	1	.2	.9	89.9
22	1	.2	.9	90.8
23	1	.2	.9	91.7
24	3	.6	2.8	94.5
25	1	.2	.9	95.4
29	1	.2	.9	96.3
31	1	.2	.9	97.2
34	1	.2	.9	98.2
71	1	.2	.9	99.1
88	1	.2	.9	100.0
Total	109	22.0	100.0	
Missing System	386	78.0		
Total	495	100.0		

4. Descriptive statistics of the h-index indicator from Web of Science

N	Valid	105
	Missing	390
Mean		3.89
Median		1.00
Std. Deviation		7.944
Range		51
Minimum		0
Maximum		51

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	49	9.9	46.7	46.7
	1	17	3.4	16.2	62.9
	2	11	2.2	10.5	73.3
	3	3	.6	2.9	76.2
	4	3	.6	2.9	79.0
	5	1	.2	1.0	80.0
	6	2	.4	1.9	81.9
	7	1	.2	1.0	82.9
	8	2	.4	1.9	84.8
	9	1	.2	1.0	85.7
	11	3	.6	2.9	88.6
	13	2	.4	1.9	90.5
	14	1	.2	1.0	91.4
	16	3	.6	2.9	94.3
	17	1	.2	1.0	95.2
	21	1	.2	1.0	96.2
	26	1	.2	1.0	97.1
	31	1	.2	1.0	98.1
	32	1	.2	1.0	99.0
	51	1	.2	1.0	100.0
	Total	105	21.2	100.0	
Missing	System	390	78.8		
Total		495	100.0		



Humboldt-Universität zu Berlin

Philosophische Fakultät I, Institut für Bibliotheks- und Informationswissenschaft

Name: Özdemir

Vorname: Tahsin

Matr.Nr.: 538959

Eidesstattliche Erklärung zur Masterarbeit

Die eingereichte PDF-Datei ist mit den Printexemplaren identisch. Ich erkläre ausdrücklich, dass es sich bei der von mir eingereichten schriftlichen Arbeit mit dem Titel:

„A Bibliometric Analysis of the Publications of the Cluster of Excellence Image Knowledge Gestaltung in the Context of Research Assessment“

um eine von mir erstmalig, selbstständig und ohne fremde Hilfe verfasste Arbeit handelt.

Ich erkläre ausdrücklich, dass ich *sämtliche* in der oben genannten Arbeit verwendeten fremden Quellen, auch aus dem Internet (einschließlich Tabellen, Grafiken u. Ä.) als solche kenntlich gemacht habe. Insbesondere bestätige ich, dass ich ausnahmslos sowohl bei wörtlich übernommenen Aussagen bzw. unverändert übernommenen Tabellen, Grafiken u. Ä. (Zitaten) als auch bei in eigenen Worten wiedergegebenen Aussagen bzw. von mir abgewandelten Tabellen, Grafiken u. Ä. anderer Autorinnen und Autoren (Paraphrasen) die Quelle angegeben habe.

Mir ist bewusst, dass Verstöße gegen die Grundsätze der Selbstständigkeit als Täuschung betrachtet und entsprechend der Prüfungsordnung und/oder der Fächerübergreifenden Satzung zur Regelung von Zulassung, Studium und Prüfung (ZSP-HU) geahndet werden.

Datum 01.07.2016